University of Dundee

University of Dundee

DOCTOR OF PHILOSOPHY

Structural studies of paired PHD finger-bromodomain chromatin-binding modules targeting epigenetic readers with chemical probes

Amato, Anastasia

Award date:
2018

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 05. May. 2023

University of Dundee

Structural studies of paired PHD finger-bromodomain chromatin-binding modules: targeting epigenetic readers with chemical probes

Anastasia Amato

Ph.D. Thesis

January 2018

Supervisor: Professor Alessio Ciulli

A mia madre e mio padre.

To my mother and my father

Acknowledgments

Firstly, I would like to thank my supervisor Prof. Alessio Ciulli, who gave me the possibility to start this adventure four years ago. Thanks for all the guidance you provided along this journey and also for the opportunity you gave me to follow my own ideas. Thanks for believing in me!

Thanks to the School of Life Science of the University of Dundee that funded this research. Thanks to all the past and current members of the Ciulli group. Everybody has contributed in some way to this experience, not only from the scientific point of view!
A special thanks to Dr. Borto, who guided my first steps into biology and was always there to support me and to…stand me! Yes, I know it was not always easy, but it was great and you handle it very well ☺
A special thanks to Dr. Lucas who greatly contributed to rescue me from a lost world. Thanks for our fruitful conversations, scientific and not!
Thanks to Teresa, sharing this experience made it a bit lighter and without you everything would have been more boring…since from the beginning. Thanks to Ester, the only one with the superpower to convince me that there is still life outside of the lab! Thanks to Wei-Wei, your friendship was a valuable discovery on the way…a bit less your 'banani' sound! Thanks to Gugui and Mattia, you were the light that suddenly appeared in Dundee ☺
Thanks to Tonia, our friendship is gonna last over the distance! Thanks to Nicola, our weekends in the lab were more enjoyable than a night in Dukes'! Thanks to Andrea, because our 'discussions' were always….productive.

Thanks to my big star, always by my side. Thanks to my mum who is the bravest woman I have ever met and she's my power!

To Marina and Rossella, sisters I've never had!

To the Joy group, because no matter where I go…our friendship lasts since ages.

Declaration

I declare that this study was carried out by myself under the supervision of Professor Alessio Ciulli. This work is original and where performed in collaboration with other people was explicitly stated in the text. This thesis has not been previously submitted for higher degree. All the references cited have been previously consulted.

Anastasia Amato

September 2017

Publication

Part of the work presented in this thesis has led to the following publication:

Chapter 1

Bortoluzzi, A., Amato , A., Lucas, X., Blank, M., Ciulli, A., *Structural basis of molecular recognition of helical histone H3 tail by PHD finger domains.* Biochem J, 2017. **474**(10): p. 1633-1651.

Summary

The use of chemical probes is a powerful tool that can help to address important biological questions.

The study presented in this thesis aims to target reader domains of chromatin-associated proteins with small molecules, in order to provide information on their ligandability, useful to develop - potent chemical probes.

This thesis work is divided in three parts.

In the first and second part it is shown how structural information obtained by the use of synthetic peptides to study the binding mode of reader domains with their natural binding partner can be combined with fragment screening to gauge future optimization of small molecules.

The first section described the disclosure of the binding mode of the H3 histone tail by the PHD zinc finger of BAZ2A. A crystal structure of the complex of BAZ2A with the H3 10-mer peptide identified a helical conformation of H3 upon binding with the PHD. This information coupled with further structural and biophysical analysis led to the identification of a subfamily of PHD characterized by an acidic patch on the helical turn, which is responsible of inducing helicity on H3 tail upon binding.

The second part of the work investigated the ligandability of the PHD zinc finger domains of BAZ2A and BAZ2B. Using a combination of biophysical techniques and X-ray crystallography it was probed that it is possible to target these reader domains. Despite the similarities of the two PHDs, comparison of the fragment-bound crystal structures of the two proteins highlighted some differences in the binding mode.

The last part of the project describes the several attempts performed in trying to elucidate the histone binding partner of the PHD-BrD tandem of the chromatin-related proteins BAZ1B and TRIM66, both involved in diseases.

List of Abbreviations

Alpha - amplified luminescent proximity homogeneous assay

ASU - asymmetric unit

BAZ1A - Bromodomain Adjacent to Zinc finger containing protein 1A

BAZ1B - Bromodomain Adjacent to Zinc finger containing protein 1B

BAZ2A - Bromodomain Adjacent to Zinc finger containing protein 2A

BAZ2B - Bromodomain Adjacent to Zinc finger containing protein 2B

BrD - bromodomain

BET - Bromodomain and Extra Terminal domain containing protein

BHC80 - synonym PHF21A

BLI - biolayer interferometry

BPTF - Bromodomain and PHD finger-containing transcription factor

BRCT - BRCA1 C terminus

BRD3 - Bromodomain-containing protein 3

BRD4 - Bromodomain-containing protein 4

BRDT - Bromodomain testis-specific protein

CHD5 - chromodomain-helicase-DNA-binding protein 5

CpG – cytosine preceding guanine

CPMG - Carr-Purcel-Meiboom-Gill

CSP - chemical shift perturbation

Da – Dalton

DCM - dichloromethane

DDT - DNA binding homeobox and Different Transcription factors domain

DMF - dimethylformamide

DMSO - dimethylsulfoxide

DNMT – DNA methyltransferase

DPF – double PHD fingers

DSF - differential scanning fluorimetry

DTT - dithiothreitol

E. coli - Escherichia coli

ECL – enhanced chemiluminescence

EZH2 – enhancer of zeste homolog 2

FBLD – fragment-based ligand discovery

Fmoc - Fluorenylmethyloxycarbonyl chloride

HAT - histone acetyl-transferases

HATU - (1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate

HBTU-(2-91H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophospahate)

HDAC - histone deacetylases

HEPES - 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HPLC - high-performance liquid chromatography

HSQC - Heteronuclear Single Quantum Coherence

HTS - high-throughput screening

JARID1A – Jumonji/ARID domain-containing protein 1A

IC_{50} - half maximal inhibitory concentration

IPTG - isopropyl-β-D-thiogalactopyranoside

ITC - isothermal titration calorimetry

K_A - association constant

Kac - acetyl lysine

K_D - dissociation constant

Kex – exchange constant

KDMs – lysine demethylase

Kme –methyl lysine

KMT – kysine methyltransferase

LB - lysogeny broth

LC/MS - liquid chromatography–mass spectrometry

LE - ligand efficiency

LH domain - Leucine-rich helical domain

MBD - Methyl-CpG-Binding Domain

MBT – malignant brain tumor

MeCP2 – methyl-CpG binding protein 2

MORF – mortality factor 4-like protein 1

MOZ – Monocytic leukemia zinc finger protein (MYST3)

MW - molecular weight

MWCO – molecular weight cut-off

MYST3 - MOZ

NMP - N-methylpyrrolidinone

NMM – N-methylmorpholine

NOE - Nuclear Overhauser Effect

NoRC - Nucleolar remodelling complex

OD_{600} – Optical density at 600 nm

PDB - Protein Data Bank

PHD - plant homeodomain

PHF21A – PHD finger protein 21A

PTM - Post-Translational Modification

Pygo2 – Pygopus homolog 2

RDMs – arginine demethylases

RFU - relative fluorescence units

RING – really interesting new gene

RMSD - root mean square deviation

RMT – arginine methyltransferase

S/B - signal to background ratio

SAM – S-adenosyl methionine cofactor

SET - Su(var)3-9, Enhancer-of-zeste and Trithorax

SGC - Structural Genomics Consortium

Sp100C – Nuclear autoantigen Sp-100

SPR – Surface Plasmon Resonance

STD - Saturation Transfer Difference

SUMO – small ubiquitin-like modifier

TAF1L - TBP-associated factor RNA polymerase 1-like

TCEP - tris(2-carboxyethyl)phosphine

TEV - Tobacco Etch Virus protease

TFA - trifluoroacetic acid

TIF1 – transcriptional intermediary factor 1

TPS - triisopropylsilane

Tm - melting temperature

TRIM24 – tripartite motif family 24

TRIM28 – tripartite motif family 28

TRIM33 – tripartite motif family 33

TRIM66 – tripartite motif family 66

Tris - trisaminomethane

TROSY – transverse relaxation-optimized spectroscopy

v/v - volume to volume ratio

WAC – (WSTF/Acf1/cbp146)

WaterLOGSY - Water-ligand observed via gradient spectroscopy

WBS- William-Beuren syndrom

WSTF – William syndrome transcription factor

Z' - Z-factor

Table of contents

CONTENTS

CHAPTER 1.
INTRODUCTION

1.1 FUNDAMENTALS OF EPIGENETICS

In eukaryotic organisms, gene expression and many cellular processes, such as cell differentiation, are under the control of epigenetic regulation. Epigenetics defines all those changes that affect the phenotype of an organism, without any alteration on the DNA sequence [1, 2]. Epigenetics can be considered as a fundamental process that monitors physiological events of the eukaryotic cell in response to environmental stimuli [3, 4]. This regulation is based on the coordination of several interdependent molecular mechanisms that modulate two distinct states of chromatin: euchromatin (associated with gene activity) and heterochromatin (associated with gene repression). Modulation of chromatin states affects many important cellular phenomena, such as transcription, replication, repair and recombination of DNA.

Chromatin is the packaging of DNA within the nucleus. The basic unit of chromatin is the nucleosome, forming the first level of compaction of DNA. The first structure of the nucleosome was solved by Luger *et al* [5] in 1997 and revealed that the nucleosome is composed of approximately 146 base pairs of DNA wrapped in a double super-helical turn around a histone octamer. Histone octamers consist of two copies of histone proteins H2A, H2B, H3 and H4. The core of the histone proteins is predominantly globular, while basic and unstructured N-terminal tails protrude away from the core of the nucleosome (figure 1.1A). Histone tails can contain many diverse post-translational modifications (PTMs) that play a key role in epigenetic regulation. PTMs are covalent and reversible modifications occurring on the side chains of specific amino acid residues (Lys, Arg, Tyr, Ser, Thr). PTMs, together with non-coding RNAs and covalent modifications on DNA e.g. DNA methylation [6], constitute a fundamental chemical platform underpinning epigenetic regulation.

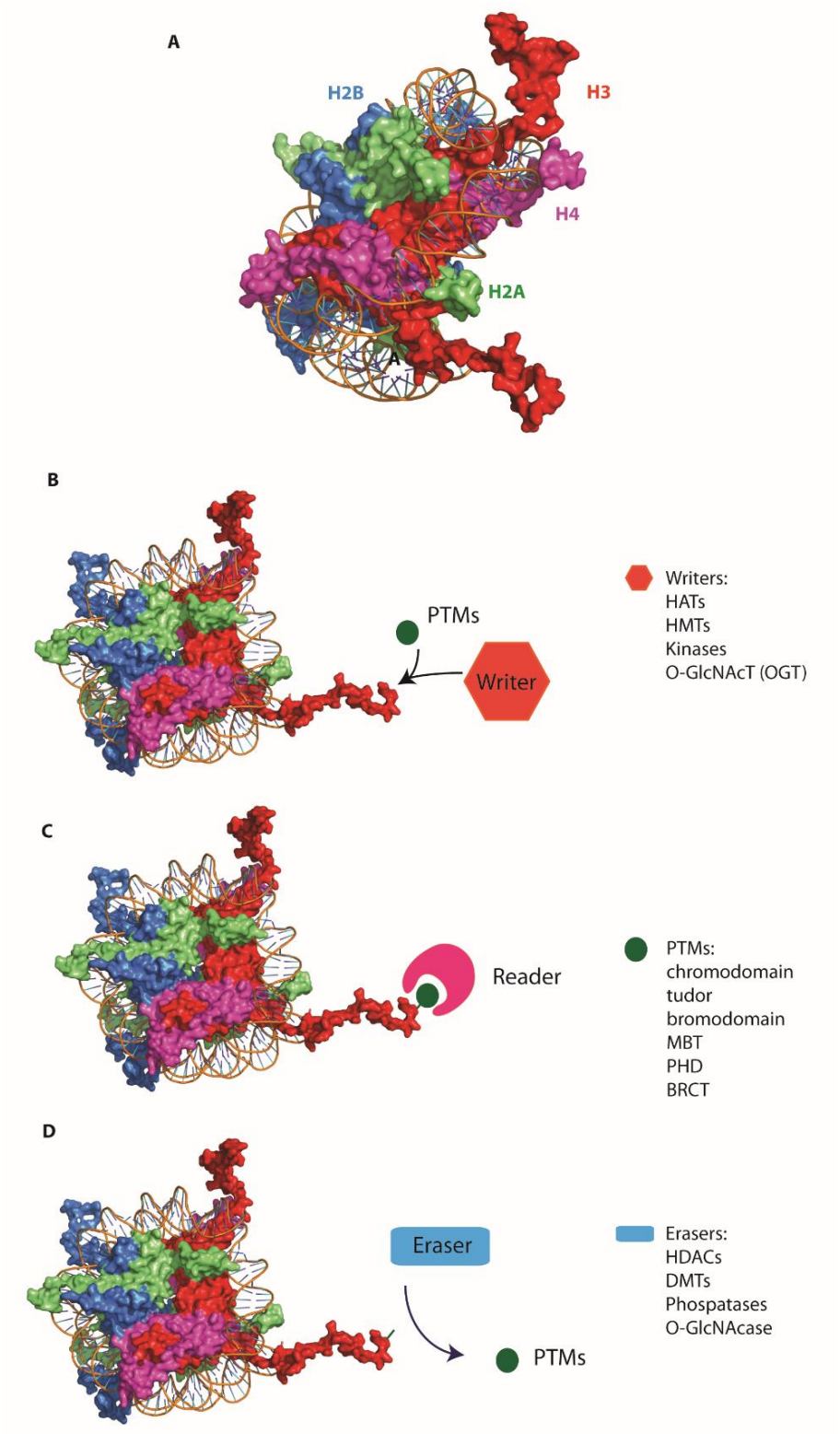


Figure 1.1. Nucleosome and histone modifications readout.

(A) Structure of the nucleosome showing the histone core and protruding histone tails as surface. In cartoon the DNA base pairs wrapped around the histone core. (PDB: 1KX5). (B) (C) and (D) respectively report a schematic representation of writers, readers and erasers of PTMs acting on the histone tails.

1.1.1 DNA methylation and reading proteins

DNA methylation constitutes one of the mechanisms which actively regulate the transcriptional states of the chromatin. Methylation is mainly occurring at the position C5 on the cytosine residue (5mC) preceding a guanine residue (CpG) [6]. The transfer of the methyl group to the pyrimidine base is catalysed by the DNA methyltransferase (DNMT) through the S-adenosyl methionine (SAM) cofactor [7].

This methylated base can be further oxidised to hydroxymethylated cytosine (5-hmC), to formylcytosine (5-fC) and to 5-carboxylcytosine (5-caC) and to cytosine again, as shown in figure 1.2.

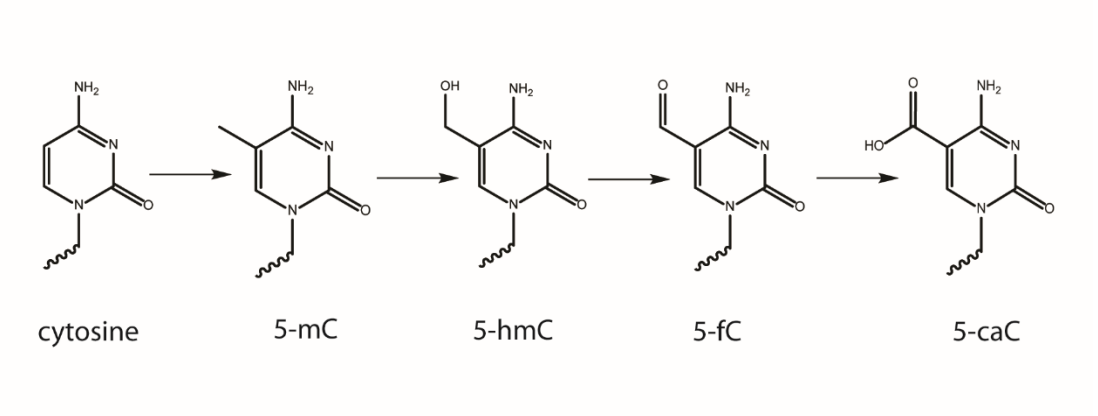


Figure 1.2. Cytosine and its derivatives.

Cytosine, methylcytosine and its oxidised derivatives.

DNA methylation has been mainly associated with gene silencing, blocking the interaction of transcription factors with the DNA while the understanding of the biological function of the oxidised derivatives is still under investigation.

Methyl-binding domain proteins, also known as MBD proteins, constitutes the major family capable of binding to modified DNA [6]. An example is represented by the MeCP2 (methyl-CpG binding protein2) which was the first MBD protein to be identified [6]. Conversely, MBD3 was the first MBD protein discovered to bind the 5-hmC motif [6]. The zinc- finger (ZF) and SET and RING-associated (SRA) family are also known to recognise 5-mC and derivatives [6].

1.1.2 Post-translational modifications (PTMs)

PTMs are chemical modification that occur on histone tails, such as acetylation, methylation, phosphorylation, ubiquitination, amongst others [4]. This class of modifications has been the main focus of this thesis work.

These modifications can be attached or removed from two antagonist classes of enzymes, respectively, called writers and eraser (figures 1.1B and 1.1D). They exert a fine-tuned epigenetic control in two ways: (i) directly, through changes on the structure of chromatin, for example, introducing repulsive interactions between histones and DNA (like acetylation or phosphorylation) or (ii) indirectly, modulating the binding of reader modules on the nucleosome.

Readers or effectors are a class of proteins capable of recognising specific pattern of sequence and structure on histone tails, containing modified or unmodified residues (figure 1.1C) [8, 9]. These readers, upon binding, can recruit enzymatic machineries able to translate this binding into a biological meaningful event [4, 10].

Two examples of PTMs that play a great role in epigenetics are acetylation and methylation. Basic science and pharmaceutical research have revealed the role of these modifications and their importance. Many cellular signalling pathways events regulated by these modifications are often associated with diseases.

Acetylation was the first modification to be identified in 1964 by the Allfrey group [11]. It is introduced by histone acetyltransferase (HATs) which transfers an acetyl group from the cofactor acetyl-CoA on the ε-amino group of the Lys side chain [12]. Histone deacetylases (HDACs) are antagonist enzymes which remove the acetyl group from Lys [12]. Addition of the acetyl moiety on Lys neutralizes the charge of the residue with remarkable impact on chromatin structure, weakening the interaction of histones with DNA, and as a result inducing a more open conformation of chromatin, becoming accessible to transcriptional machineries [13]. This can explain how acetylation is often uniquely associated with high transcriptional activity [13].

Another important modification occurring on both Arg and Lys residues is methylation[14]. It is a dynamic modification [15], which is regulated by RMTs (arginine methyltransferase) and KMTs (lysine methyltransferases) antagonistic to the R(K)DMs (demethylases) [16 17-23]. These PTMs are simultaneously found in association either with repressive or activating states, according to the level of methylation and to the position of the residues on the histone. Methylation on Lys can occur as mono-, di- or tri-methylation. On the Arg, methylation can occur as mono- or di-methylation in a symmetric (on both nitrogens of the guanidinium group) or asymmetric (only on one nitrogen of the guanidinium) manner [14, 24]. Arg methylation does not affect the overall charge of the residue but it can modulate the types of interactions that the side chain can establish. A remarkable example of how these marks can have different biological outcomes is mono-methylated K4 (H3K4me) which is often associated with gene activation in combination with H3K27ac [4], while H3K9me3 and H3K27me3 are examples of methylation found associated to repressive states in heterochromatin [4, 25].

Phosphorylation is another modification occurring on serine, threonine and tyrosine. It is found on histone and non-histone proteins on which it is dynamically regulated by kinases and phosphatases, two antagonistic enzymes. This modification adds a negative charge on the histone with consequent effects on the chromatin structure. It is often found associated with other modifications with which it acts in cross-talk [26].

β-N-acetylglucosamine (O-GlcNAc) has also been recently identified as a novel epigenetic mark of the histone code [27]. This modification, occurring on the side chain of serine or threonine, is mainly involved in regulatory processes of non-histone proteins [28]. Work from Sakabe *et al.*, in 2010, identified for the first time glycosylation marks on histone tails [27]. From that seminal study, this PTM has emerged as involved in gene expression and regulation of chromatin remodelling complexes [29].

Figure 1.3 depicts a schematic illustration of the major PTMs identified on the N-terminal histone tails of the nucleosome.

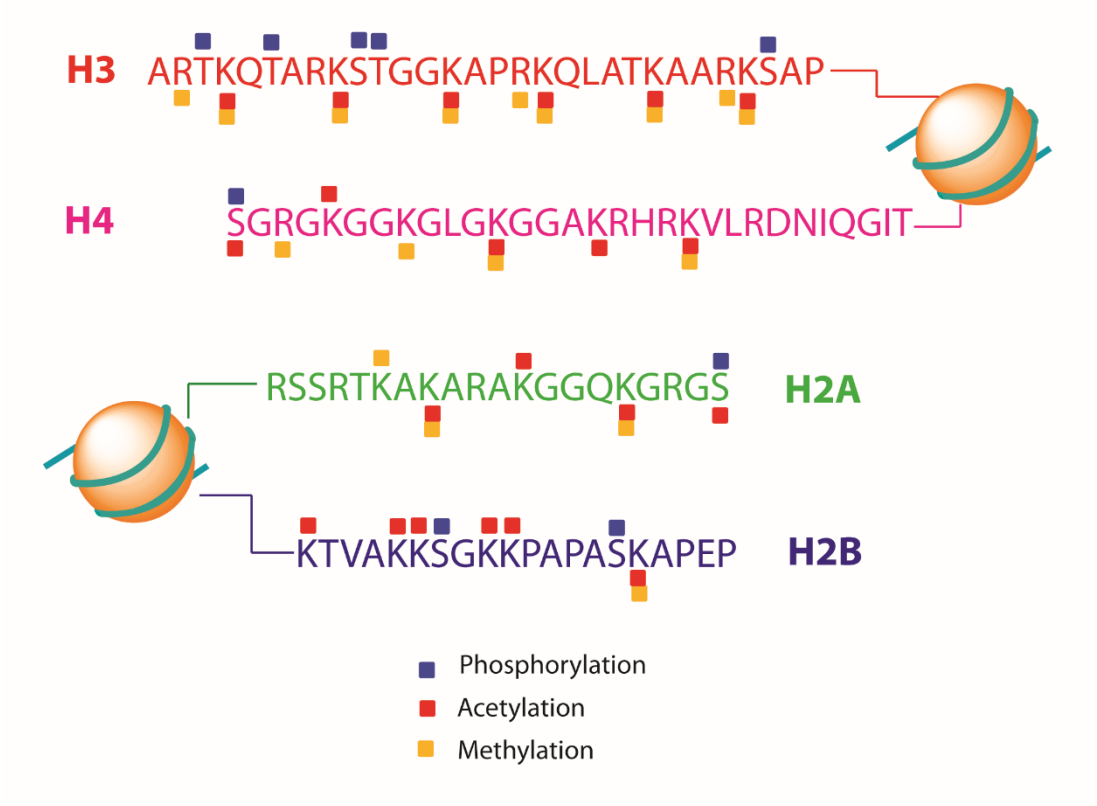


Figure 1.3. Simplified illustration showing PTMs on the four histone tails protruding from the nucleosome.

1.1.3 PTMs reader domains

Reader domains are protein modules able to bind specific sequences and marks on histone tails [8, 9] with the final aim to translate this binding in a biological meaningful event.

Table 1.2 (adapted from Taverna *et al*. review [8]) groups the different known reader domains and the relative PTM recognized.

READER	PTM marks
Bromodomains	Kac
Royal family :	
chromodomain	H3K9me2/3, H3K27me2/3
double chromodomain	H3K4me1/2/3
chromo barrel	H3K36me2/3
Tudor	(Rme2s)
Double/tandem tudor	H3K4me3, H4K20me3 H4K20me1/2, (Kme2)
MBT	H3K20me1/2, H1K26me1/2 H3K4me1, H3K9me1/2
PHD finger	H3K4me1/2/3, H3K4 H3K9me3, H3K36me3 H3K14 acylations
WD40 repeat	H3R2/K4me2
14-3-3	H3S10ph, H3S28ph
BRCT	H2A.X-S139ph

Table 1.1. Summary of reader domains and associated PTMs [8].

Summary of the principle reader domains (left column) and the marks recognized (right column).

One of the most studied reader domains is the bromodomain, which recognizes Kac marks on histone tails as well as other proteins [30, 31]. Bromodomains play a key role in many protein-protein interactions in cell [30], thereby motivating intense interest in this domain [32]. Methyllysine reader domains have also received attention in the last few

decades. Amongst these readers are: PHD zinc fingers, chromodomain, chromo barrel, tudor domain and MBT (malignant brain tumor) domain [8, 9]. The specificity of these modules is dictated by peculiar features of the histone binding pocket. According to the level of methylation of recognized residues, the hydrophobicity of the binding pocket is increased [8].

Much less is known about recognition of phosphorylated marks. Indeed to date only a couple of examples of effector modules have been reported in literature: the 14-3-3 scaffold and the BRCT domain of MDC1, which, respectively, recognizes H3S10ph [33] and S139ph on H2A.X [34].

Herein, I will examine in greater details two of the reader domains that were the subject of investigation of this PhD thesis – the bromodomain and the PHD finger domain.

1.1.3.1 Bromodomain

The bromodomain specifically recognises ε-N-acetylated lysine. It is a highly structurally conserved domain despite the relatively low sequence identity. This effector module was identified for the first time in 1992 in the *Brahma* gene of the *Drosophila melanogaster* [35]. It is composed of a central hydrophobic pocket determined by a left-handed bundle of four α-helices (designated as α Z, α A, αB and αC) linked with loops of variable length (named as ZA and BC loops) [8, 36, 37]. A highly conserved asparagine residue is responsible for the coordination of the bound Kac together with a network of structurally conserved water molecules [36, 37]. In the human genome, 61 different bromodomain modules have been identified, belonging to 46 different chromatin-associated proteins [37]. A phylogenetic tree (figure 1.4B) groups the different bromodomain in seven subfamilies, according to structural similarities [37]. These proteins are involved in different transcription networks, and are often associated with disease pathways, for example, cancer, inflammatory response or viral infections [30]. For these reasons, bromodomains have been identified as potential therapeutic targets [30] and several inhibitors have been generated to disrupt protein-protein interactions involving these effectors [36 38-41].

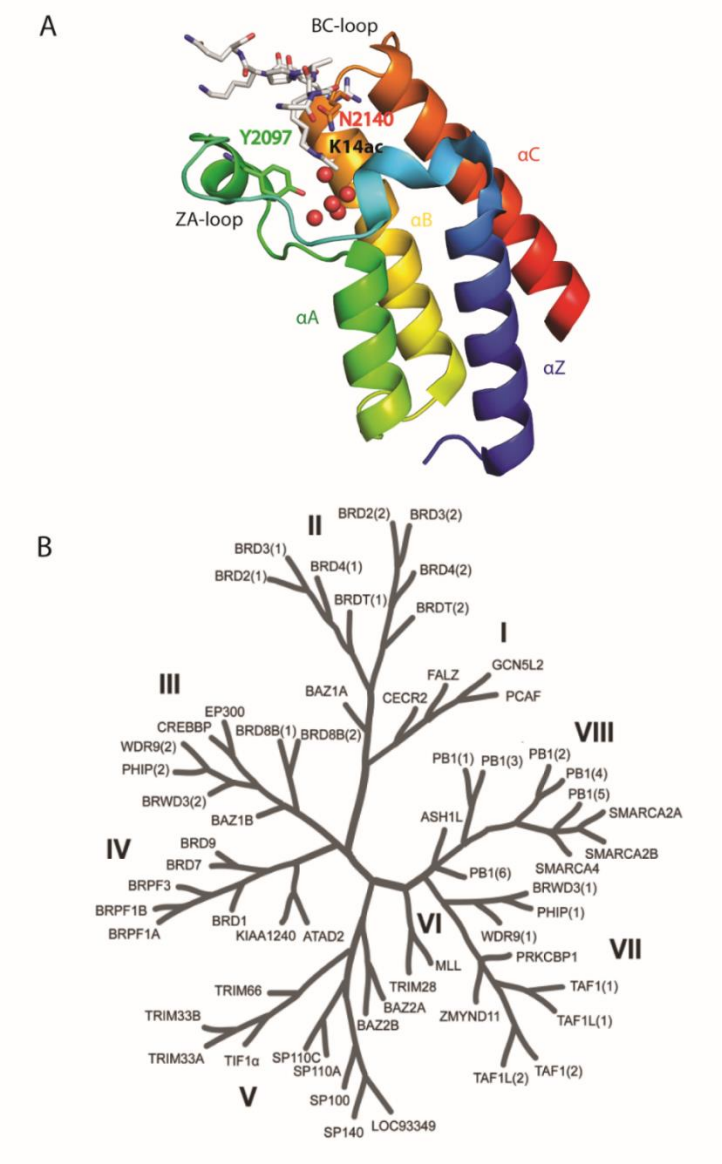


Figure 1.4. Bromodomain structure and phylogenetic tree.

A- Cartoon structure of the bromodomain of BAZ2B in complex with K14ac in sticks (PDB:4QC1). The five conserved waters of the pocket are reported as sphere and the two conserved residues as stick.

B- Phylogenetic tree of the bromodomain reporting eight different subfamilies (I, II. III etc.). Classification is based on structure similarity. Figure was adapted from Filippakopoulos *et al*. [37].

1.1.3.2 PHD zinc fingers

Plant Homeodomain (PHD) zinc finger is a small domain constituted of 50-80 amino acids characterized by the highly conserved motif of *Cys3-His-Cys4* that coordinates two zinc

ions in an interleaved fashion, as shown in figure 1.5. It takes the name from the plant homeodomain of *Arabidopsis thaliana* where was identified for the first time in 1993 [42].

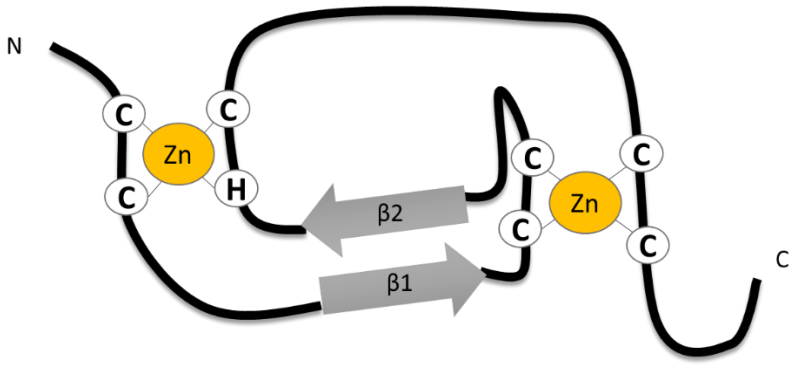


Figure 1.5. PHD zinc finger fold.

Schematic view of the PHD zinc finger fold characterized by the double β-strands responsible of the interaction with the histone and the conserved motif Cys4-His-Cys3 that coordinates two zinc ions. (Figure adapted from Musselman *et al*. [43]).

More than 120 sequences have been annotated as PHD zinc fingers. Figure 1.6 illustrates the phylogenetic tree grouping the different domains according to their sequence similarities.

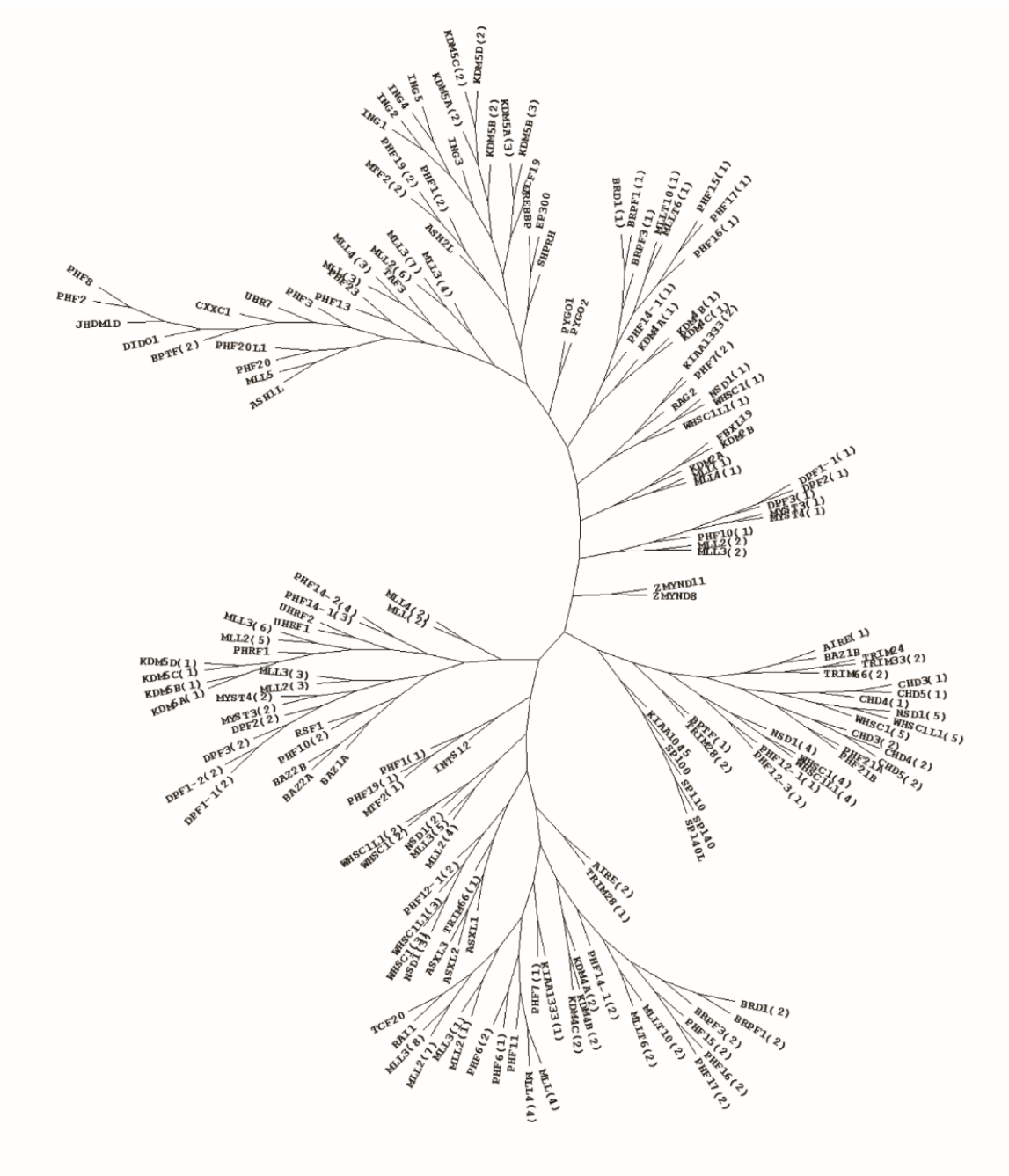


Figure 1.6. Phylogenetic tree of PHD obtained from the *ChromoHub, GSK*.[44].

Phylogenetic tree is obtained from sequence similarities upon alignment.

In contrast to other reader domains that selectively bind specific modifications, PHD fingers have been shown to be much more diverse with respect to the choice of their binding partner. This diversity can be related to the multiple biological functions in which they are involved. Some PHD fingers can specifically recognise unmodified regions of the N-terminal tail carrying H3K4 with no modifications (figure 1.7 C) [45-49]. Other PHDs specifically recognise H3K4me3 or H3K9me3 (figure 1.7 A and 1.7 B) [50, 51]. One PHD

finger displayed preference of binding for H3K36me3 [52]. Another subset of PHDs, recently identified, reported binding toward different acylation states on position K14 [53]; peculiarity of the pocket of this specific subset is the presence of a conserved Gly in the first β-strand that replaces the bulky residue found in other PHDs [53].

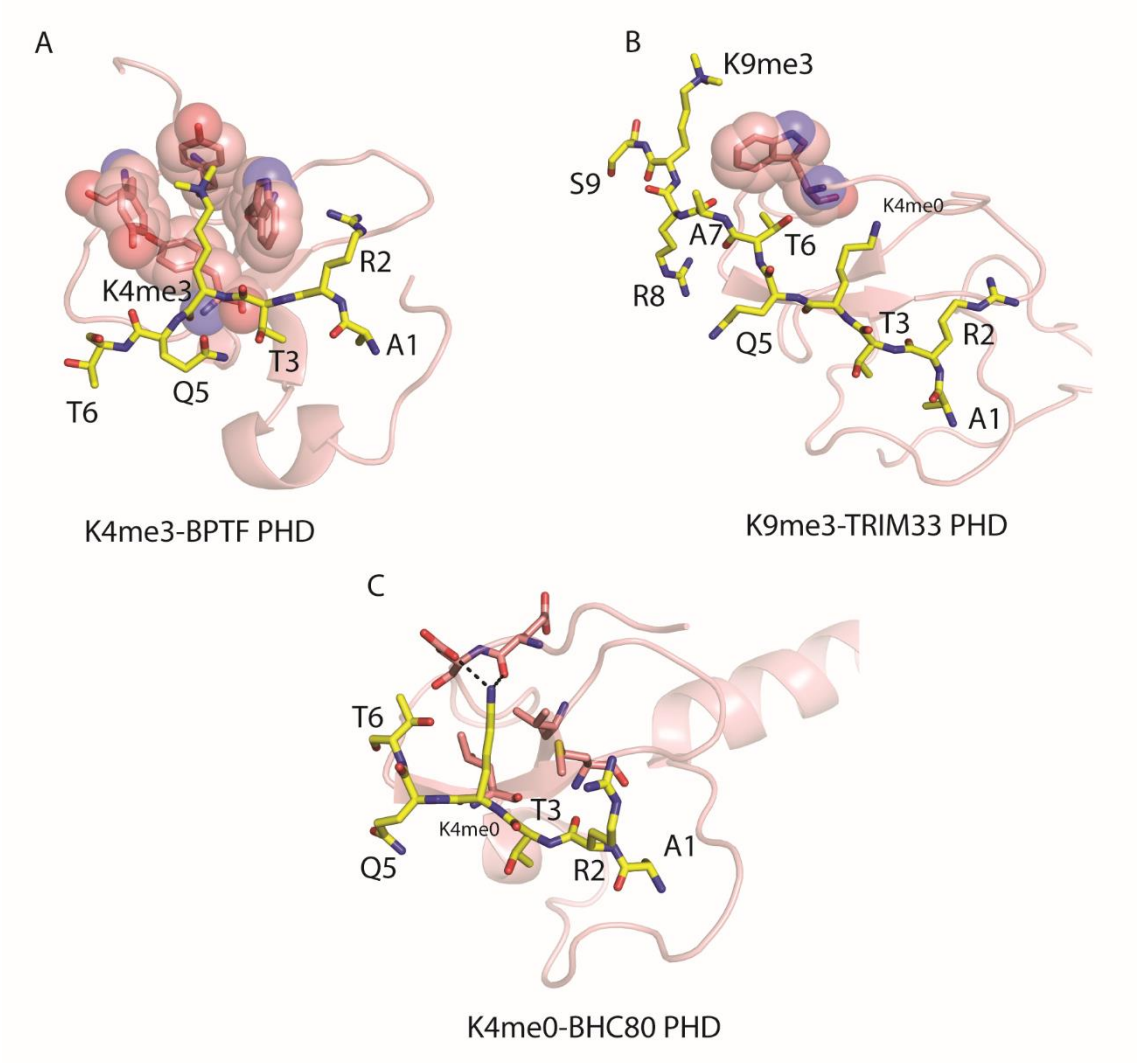


Figure 1.7. A-C. Structural basis on the recognition mode of PHD fingers.

(A) Crystal structure of PHD of BPTF in complex with H3K4me3 (PDB: 2F6J). (B) Crystal structure of PHD of BHC80 in complex with H3K4me0 (PDB: 2PUY). (C) Crystal structure of the PHD of TRIM33 in complex with H3K4me0K9me3 (PDB: 3U5N). PHD is represented as surface and histone peptides as yellow sticks with nitrogen and oxygen, respectively, in blue and red.

PHD domains involved in H3K4me3 recognition contain a conserved aromatic cage (figure 1.7 A). PHD fingers that lack such a signature instead recognize unmethylated K4 (figure 1.7 B-C). At low methylation states, hydrogen bond acceptors facilitate the

interaction with K4. In contrast, at increasing methylation states, aromatic residues such as F,Y or W play an important role in the molecular recognition, due to the formation of stable cation-π interactions with the charged methylated primary amine group [54].

PHD zinc fingers have recently been investigated as potential targets for therapeutic intervention, as several biological and genetic studies have linked many proteins containing PHD finger domains to human diseases [43, 55]. However, in contrast to bromodomains, not many studies have addressed the ligandability of these reader domains. To date, there are only two papers reporting inhibitors of PHD zinc fingers. In the work presented from Wagner *et al*., a chemical scaffold was identified that disrupted the interaction of the PHD finger of JARID1A with a H3K4me3 peptide [56]. Mutagenesis experiments and modelling studies suggested a hypothetical region of binding for the small molecule, however no structural or biophysical data were provided to confirm specific, reversible protein-ligand binding [56]. In 2014, Miller *et al*., using *in silico* and NMR screening, identified several fragments able to bind the PHD finger of the protein Pygo. Amongst them, the best compound led to the first fragment-bound structure (figure 1.8). A K_D of 2.5 mM was estimated by CSPs [57]. This work can be considered pioneering in terms of fragment screening to identify potential starting points toward the development of small-molecule inhibitors of the PHD-histone interaction.

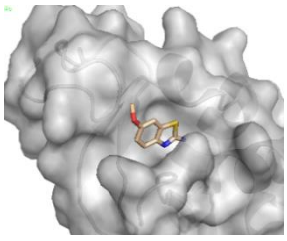


Figure 1.8 First fragment bound structure of a PHD finger.

Crystal structure of the PHD of Pygo2 (grey surface) in complex with a chemical fragment CF4 (in stick). PDB: 4UP5.[57]

1.1.4 Combinatorial readout by tandem modules

Epigenetic proteins often contain more than a single reader/writer/eraser domain. Genome-wide sequence analyses have led to the identification of a frequent

concurrence of multiple reader domain within the same protein. This observation led to the hypothesis of a possible combinatorial readout associated with multivalent binding to impart sequence specificity, and concomitant recognition of distinct post-translational modifications [8, 9]. Adjacent reader domains could act in cross-talk by binding different combinations of marks on the histone tails. Examples are the PHD domains from MOZ, DPF2 or MYST3 that are able to recognize unmodified lysine on one PHD and acetylated K14ac on the adjacent PHD [8, 58]. Also the PHD fingers of CHD5 constitute an example of multivalent engagement of two adjacent PHDs both binding to unmodified H3 tails [59].

Thus, combination of modifications can actually act as a single unit, a 'modification cassette' [60], able to be recruited by different effectors in a concerted manner. Combinatorial modifications are not only found on the same histone tail but they can also play in cross-talk on different histones belonging to the same nucleosome or on distinct nucleosomes [61]. Picture 1.9 illustrates the different possibilities offered.

The relevance of this multivalent mode of action is in the opportunity of modulating the same system in different way, in order to explicate distinct biological outcomes.

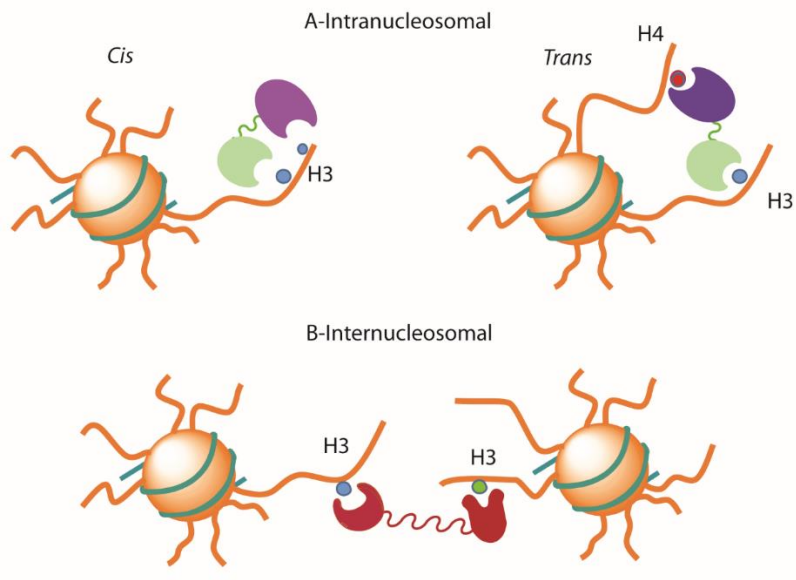


Figure 1.9. Multivalent engagement by reader domains.

(A) Examples of adjacent reader domain recognizing in *cis* and *trans* histone tail situated on the same nucleosome. (B) Examples of adjacent readers recognising histone tail located on different nucleosome.

An example of combinatorial action in *cis* is offered by the PHD-Bromodomain of TRIM33 and TRIM24 which respectively recognize H3K9me3K18ac [51] and H3K4me0K23ac [62]. Both these tandem modules contain reader domains connected by a small loop that facilitate mutual interaction of the two reader domains and recognition of the modification cassette on the same histone tail H3. The PHD-Bromdomain of BPTF, instead, is a representative example of recognition in *trans*. In this protein, the two effectors are connected by a α-helical linker which keeps them far apart, so that the PHD binds the H3K4me3 mark, while the bromodomain recognizes H4K16ac on a separate histone tail [50, 63].

Multivalent engagement can be explained by introducing the concept of cooperativity [64]. The recognition of specific marks is not an isolated event involving only a single residue but is rather influenced by adjacent residues or PTMs that can fine-tune (positively or negatively) the binding affinity. For example, the Sp100C PHD, which recognizes unmethylated H3, showed higher affinity for the histone tail in presence of phosphorylated S10 [65]. For the PHD-Bromodomain tandem of TRIM24 and TRIM33, it was observed that the longer length of the peptide is critical to achieve binding affinity and specificity [51, 62]. By contrast, a negative tuning is observed for the binding of HP1 to H3K9me2/3 which is disrupted in presence of H3pS10 [60]. This antagonistic mechanism is also known as 'binary switch' [60].

Study of the cooperative effect on reader tandem modules have the advantage to be more informative than studies with a single domain, allowing better understanding of the mechanism of engagement of the histone tails by chromatin-associated proteins [64].

1.2 TARGETING THE EPIGENOME

Despite progress achieved in this last decades in the understanding of the mechanisms underpinning epigenetic regulation, several questions remain to be addressed. In this context, a valuable approach can be the development of chemical probes able to bind with high affinity and selectivity a particular reader domain. Such probes could be used as tools to dissect the role and explore the function of a protein or of its individual domains [66 67]. The advantage of using chemical inhibitors is in the possibility to edit a specific function of the protein [68] instead of blocking the expression of the whole protein as it is the case with gene knockout or knockdown approaches. An example of recent successful application of chemical probes to biological discovery is the work presented from Sdelci *et al* that used small molecules to reveal how TAF1 and BRD4 synergistically control proliferation of cancer cells [69].

There is strong evidence of the involvement of the epigenome in the establishment of important human diseases such as cancer, inflammation, diabetes, neurological disorders, amongst others [43, 55, 70-72]. Several studies have documented clinical relevance of reader domains, electing them as potential therapeutic targets. Many chromatin-associated proteins have been shown to have medical relevance and amongst these, chromatin-modifying enzymes have proven to be enticing drug targets which has led to five drugs now approved by the FDA (Food and drug administration) [73]. In the last decades a large number of human cancers has been often associated with DNA methylation [74]. For instance, mutations occurring on the IDH1/2 (isocitrate dehydrogenases) proteins, which inhibit the activity of histone demethylases and DNA methylases, lead to a hypermethylation of DNA and histone which is characteristic of the phenotype of the acute myeloid leukemia (AML) [74]. Decitabine, marketed as Dacogen, is an example of epigenetic drug proven to be effective against AML. Decitabine is a cytosine analogue which can be incorporated in DNA and RNA, reducing the level of methylation through inhibition of the methyltransferases [75]. Also Histone deacetylases (HDACs) have been strongly targeted in the last decades since their overexpression has often been linked to certain types of human cancer [76]. Vorinostat, commercially known as Zolinza [77], represents another example of epigenetic drug

already on the market. It acts as inhibitor of the HDACs of class I, II and IV which contain a zinc in the active site. Its mechanism of action is based on the chelation of the catalytic zinc [78] and it was proven to be mainly effective against the cutaneous T cell lymphoma [79].

Amongst reader domains, bromodomains represent the most targeted domains. About 19 lead compounds targeting bromodomains are currently being evaluated in clinical trials, notably inhibitors of the BET bromodomains family [80], as for example I-BET762 [81] and (+)-JQ(1) [82]. Amongst other reader domains, potent inhibitors have also been discovered for methyllysine binders like the MBT (malignant brain tumor) domain, chromodomains and tandem tudor domains [83].

In light of these developments, drugging the epigenome can be perceived as a great opportunity to: 1) provide chemical probes as useful tools to advance in the understanding of the humane interactome [84] and 2) expand the druggable human genome, an important goal for drug discovery today.

1.2.1 The challenge of targeting protein-protein interactions of epigenetic readers

Targeting epigenetic reader domains requires addressing protein-protein interactions (PPIs) [84]. The challenge in targeting PPIs with small molecules emerges from the observation that the surface involved in the binding between two proteins is much larger when compared to the area covered by small molecules. It is been estimated that on average the surface area between two interacting proteins is ~ 1500–3000 $Å^2$ [85] while the size of a small molecule is ~ 300–1000 $Å^2$ [86]. The region of interaction between two proteins is generally a flat and shallow surface that tends to lack grooves and deep pockets which are typically found on proteins that bind natural small molecules (such as kinase or GPCR receptors) [87]. Indeed, in drug discovery approaches, targeting kinases or GPCR families, the chemical scaffold of endogenous ligands is often used as starting point to develop synthetic, artificial antagonist.

Albeit PPIs seem to involve a large surface area, it has been shown that, in many cases, only few residues are de facto contributing to the energy of binding. These residues are

called hot spots of interaction [87] [88, 89]. These hotspots are generally clustered at the centre of the interface and can cover an area comparable to the size of a small molecule. .

PPI can be classified in three categories, according to the kind of epitope which they recognise:

(i) primary, if one side of the interface can associated with a linear protein sequence, as the case of PTMs recognition;

(ii) secondary, if they recognise epitope with one secondary structure;

(iii) tertiary, if they recognise a more complex epitope characterized by a tertiary structure organization.

Peptides are useful tools in structural biology to gain insight the molecular recognition of PPIs [87, 90]. In the case of primary or secondary PPIs, peptides can simulate the protein surface, resembling the amino acid sequence of the natural binding partner. Their usage, coupled with biophysical techniques and mutagenesis experiments, can lead to the identification of the hotspots of interaction. The identification of hotspots may therefore assist in the generation of novel small molecules as specific disruptors of the PPIs.

1.2.1.1 Targeting the epigenome: druggability or ligandability studies?

There is often confusion in the usage of the terms “druggability” and “ligandability”, which are erroneously used as interchangeable terms. They actually refer to two different concepts. The term ligandability generally describes the possibility to identify a chemical scaffold able to bind a designated target with a certain affinity [91]. If the targeted binding site is involved in the binding of a natural binding partner e.g. another protein or a small-molecule substrate or cofactor, then a ligandable binding site implies the propensity to generate suitable ligands of good chemical properties, that can effectively compete with the natural binding partner. The concept of ligandability can be visualized adopting the lock and key hypothesis, introduced by E. Fisher and L. Pauling, and generally used to describe the mode of action of the enzymes [92] . The lock is the target protein with its pocket/surface and the keys are the different molecules that can potentially bind the pocket. Only one key will be able to enter the lock –

exemplifying how only one molecule will bind specifically the target. Target ligandability provides a first necessary (but not sufficient) step towards target druggability. Target ligandability is the concept to which I will mainly refer in this thesis work.

Druggability is a notion that engages multidisciplinary concepts because it implies that a suitable small molecule also satisfies certain pharmacodynamic and pharmacokinetic parameters and criteria that make it compliant and acceptable as potential drug i.e. drug-like [93]. Moreover, such molecule should be able to abolish or at least reduce the aberrant effects of the disease pathway in which the target is involved, in order to generate a therapeutic effect for the patient [94].

1.2.1.2 How to target protein-protein interactions: Fragment-based lead discovery (FBLD)

In the last years several approaches have been developed and applied to target protein-protein interactions. The established high-throughput screening of a large library of compounds (HTS) has been rivalled by the advent of fragment-based lead discovery (FBLD). The paternity of the concepts and theory behind fragment-based approaches is ascribed to Jencks' work in the far 1981, on the additivity of binding energies of small molecules [95]. His work paved the way to the idea that smaller component (fragments) could be productively assembled and optimized to produce larger molecules capable of better quality of binding and increased binding affinity. The difference in concept between HTS and fragment approach is illustrated in figure 1.10.

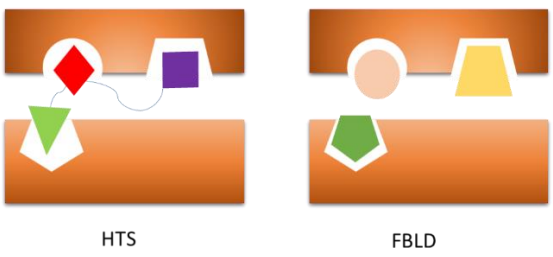


Figure 1.10. Schematic differences between HTS and FBDD approach.

A- Poor quality of protein-ligand interactions in HTS. B- The small size of the fragments allow each molecule to bind specific region of the interface in an optimal way.

In HTS screening large molecules typically make non-optimized interactions due to non-specific binding. In contrast, in the fragment-based approach, each fragment can specifically target a small region of the protein. The potential of the FBLD approach is the possibility of combining in different ways the best hits obtained from screening (merging, linking or growing them) in order to develop a bigger scaffold with optimized interactions toward the target [96]. Optimization steps of the fragments are generally addressed using structural knowledge on the mode of binding of the fragment itself. X-ray crystallography and protein-observed NMR are leading techniques to provide these information.

Fragment screening can be performed *in silico* or involving a cascade of complementary biophysical techniques which provide high sensitivity in detection of low affinities. Estimation of the affinities of the ligand is a key step to help to rank the fragments through the determination of either a K_D or IC_{50} and the relative ligand efficiency (LE) [97]. The concept of ligand efficiency is often used in medicinal chemistry to capture in a single metric how much binding affinity is contributed by each atom of the molecule. This metric provides a useful criteria for selecting the best fragments to pursue for further optimization, and can be much more informative than binding affinity alone. LE is defined by a simple equation (1) which divides the free energy of binding by the number of non-hydrogen atoms. Fragments with LE higher than 0.3 kcal $\times$ mol^{-1} per atom are generally considered efficient binders, albeit LE of ligands can closely depend on the ligandability of the target-binding site. For examples, ATP binding sites of protein kinases readily yield fragment binders with LE > 0.4, while flat and featureless PPIs often yield hits with LE ~ 0.2.

$$LE = \frac{\Delta G}{NHA} \qquad (1)$$

Other considerations in addition to binding affinity and LE should be kept in account when selecting fragment hits. For example, knowledge on the binding mode of the fragment and chemical tractability of the scaffold for future optimization are other important parameters to consider.

1.3 TARGETING PHD ZINC FINGER AND BROMODOMAIN TANDEM MODULE

In this thesis project I was interested in the study of PHD - Bromodomain tandem module of proteins involved in important biological pathways and linked to diseases: BAZ proteins and TRIM66. In the next subchapter I will give a brief introduction about these proteins and why it was decided to study them.

1.3.1 BAZ family: bromodomain adjacent to zinc finger

In 2000, the Shimane group in Japan identified a new family of four proteins, containing bromodomain adjacent to zinc finger, which was named as BAZ family [98] (figure 1.11). Members of this family all contain a PHD zinc finger and bromodomain at the C-terminus in addition to a DDT domain (*D*NA binding homeobox and *d*ifferent *t*ranscription factors). Additionally, BAZ2 proteins specifically contain a MBD (*m*ethyl-CpG-*b*inding *d*omain) motif, also known as TAM domain. The WAC motif was a trait identified only in BAZ1 proteins and is required for DNA binding [99].

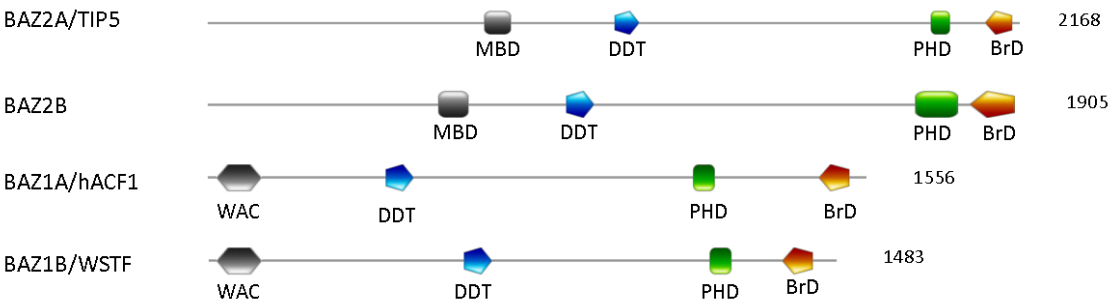


Figure 1.11. BAZ family.

Proteins belonging to the BAZ family with representative locations of conserved motif sequences. Most sequence variation is seen at the amino terminus where BAZ1A, BAZ1B have a WAC motif while BAZ2A, BAZ2B lack that [45]. Picture was obtained using Prosite MyDomain [100]

.

In this project, I focused my attention on three proteins of the family: BAZ2A, BAZ2B and BAZ1B, which are described in details below

1.3.1.1 BAZ2A (or TIP5)

BAZ2A (also known as TIP5 -Transcription termination factor I-interacting protein 5) is one of the more studied members of the BAZ family. It has been previously identified as part of NoRC (nucleolar remodelling complex) which uses BAZ2A as scaffold to bind to chromatin and subsequently recruit other components of the complex [101, 102] . The role of BAZ2A as part of the NoRC complex is essential to assemble and settle the heterochromatin state. BAZ2A was found overexpressed in prostate cancer interacting with the methyltransferase protein EZH2 to induce aberrant silencing in cancer [103]. Although its mechanism of action has not been fully elucidated, BAZ2A has been identified has a possible biomarker for diagnosis of the prostate cancer [103]. These discoveries point towards a clinical relevance for this protein. This motivated our interest in targeting BAZ2A. The presence of the PHD-BrD tandem modules on the protein skeleton makes this protein also an attractive model system for molecular recognition studies of multivalent histone binding.

1.3.1.2 BAZ2B

BAZ2B is the less studied protein of the family. Indeed, its biological function is unknown and speculations are only arising considering the high similarity with BAZ2A protein, especially in the region of the PHD-Brd tandem (60% identity) [45]. To date, the only relevant knowledge about BAZ2B function is its association with a doubling of sudden cardiac death (SCD) risk [104]. The lack of a functional role for this protein provides the motivation to drive the development of chemical probes useful to explore all the possible biological functions and interactions of BAZ2B.

1.3.1.3 BAZ1B (or WSTF)

BAZ1B is an atypical tyrosine-kinase protein [105], also known as WSTF protein. It is considered a versatile chromatin-modifying factor since it has been identified in at least three different complexes, all involved in distinct chromatin –related processes [106]. The gene coding for this protein is part of a sequence that if deleted leads to the William-

Beuren syndrome (WBS) [107, 108]. Insufficiency of BAZ1B protein could be the reason behind several anomalies typical of the WBS [109]. Expression of this protein is found in almost all the tissues, including foetal tissues [106].

How the binding of this protein is occurring at chromatin level is still unclear. Studies on the PHD-Bromodomain module can be a good starting point to answer these questions and reach a more complete understanding of its function and involvement in diseases. In light of this, to provide a chemical probe able to target and modulate this protein could be a strategic approach to unveil the biology of BAZ1B.

1.3.2 TRIM: Tripartite motif family

The tripartite motif family (TRIM) includes a large number of proteins all characterized by the presence of a RING domain [110], 1 or 2 B-box motifs [111] and an adjacent coiled-coil motif [112] at the N-terminal region. The majority of these proteins are found to be linked to cancer either in oncogenic or tumour suppressor activity. Within this family, which includes about 70 members, proteins are differentiated in 12 subgroups, based on their variations at the C-terminus [113]. Transcriptional intermediary factor 1 (TIF1) family is one of the subgroups of the TRIM family and includes four proteins characterized by a PHD-BrD tandem module at the C-terminus (figure 1.12).

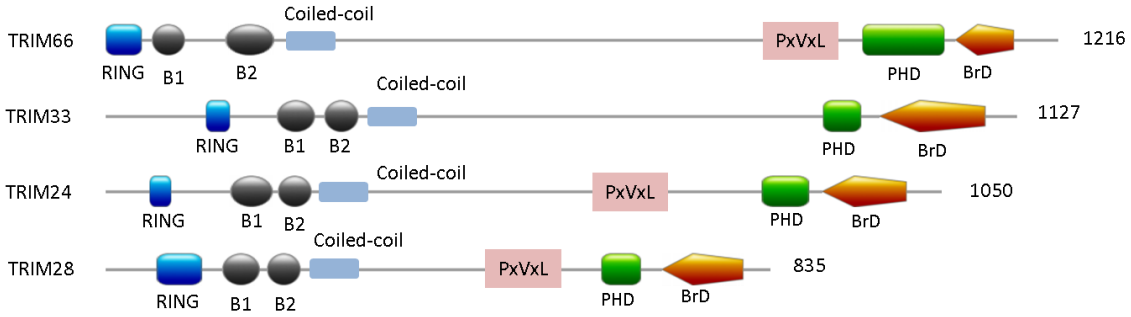


Figure 1.12. TIF1 family.

Schematic illustration of the domains constituting the four proteins belonging to the TIF1 family. Figure was obtained using PROSITE MyDomains [100].

In this project I focused the attention on TRIM66, which is the less characterized protein of the TIF1 family.

1.3.2.1 TRIM66 (or TIF1δ)

Trim66, also known as TIF1δ, was the last component of the family to be identified, in 2004 [114]. While other members of this subgroup have been studied in greater detail, very little is known about TRIM66. This protein was discovered as a novel HP1-interacting member of the transcriptional intermediary factor 1 (TIF1) family [114]; indeed TRIM66 features a PxVxL motif needed for the binding to HP1, as TRIM24 and TRIM28. It shares conserved domains with the other member of the family (figure 1.12).
Expression of TRIM66 was observed only in testis, differently from the other members of the family which are ubiquitously expressed in all tissues. Scientific evidence suggested that this protein could be part of transcriptional complexes that regulate a repressive state of chromatin [114]. TRIM66 was recently linked to lung cancer progression [115] and it is highly overexpressed in osteosarcoma cells [116].
This information highlights the need to improve our scientific knowledge on TRIM66 as it seems to play a key role in cancer. This motivated the study on the potential interaction of TRIM66 with chromatin *via* its PHD-BrD module. The final aim is to provide structural knowledge to underpin and drive future targeting of TRIM66 with small molecules.

1.4 BIOPHYSICAL TECHNIQUES

The following section will briefly introduce the main biophysical and biochemical techniques used in this thesis work.

1.4.1 Nuclear magnetic resonance (NMR)

Nuclear magnetic resonance spectroscopy is a powerful technique extensively applied in structural biology to study biomolecules interactions in solution. Two formats can be used: (i) protein-observed NMR and (ii) ligand-observed NMR.

1.4.1.1 Protein-based NMR (^{1}H-^{15}N HSQC)

In protein-observed NMR, observable signals belong to the protein which therefore requires to be labelled with isotopes as ^{15}N or ^{13}C and ^{2}H [117].

(^{1}H-^{15}N)-HSQC (Heteronuclear single quantum coherence spectroscopy) is a widely used experiment to study the interaction of proteins with either biomolecules or small molecules [118]. It requires only ^{15}N labelling. In the HSQC spectrum, peaks are generated by the transfer of magnetization from the hydrogen of the peptide bond to the ^{15}N nuclei through J-coupling. After a time delay (T1), the magnetization is transferred back to the hydrogen where is detected. Each amino acid of the protein can univocally generate a peak on the HSQC spectrum, with the exception of Prolines and the N-terminal residue (that contains NH_3^+ which exchanges with the solvent and is not observable). The assignment of each peak to a residue can be performed using a set of 3D experiments that establish sequential connectivities between residues, therefore the resonances can be linked to distinct residues and sequentially assigned.

Temperature, buffer and pH of the solution can significantly affect the position and quality of the peaks observed. Also addition of a ligand to the protein can affect the position of the protein resonances generating a shift of the peaks involved in the binding with the ligand. Shifts can be quantified using the equation (2).

$$\Delta\delta\text{weighted} = \sqrt{|\Delta\delta\text{H}|^2 + |\Delta\delta\text{N}|^2 * 0.15} \qquad (2)$$

Where ΔδH and ΔδN are respectively the shift on H and N scale.

Analysis of the shifts can lead to the mapping of the binding site on the protein. This procedure is called chemical shift perturbation (CSPs) [118, 119]. CSPs performed with increasing concentration of ligands can allow estimation of a dissociation constant (K_D), using the equation (3).

$$\Delta\delta \text{ obs} = \Delta\delta \max \frac{\left\{Kd+Pt+Lt-\sqrt{(Kd+Pt+Lt)^2-4Pt+Lt}\right\}}{2[P]t}$$

[P]t and [L]t are, respectively, the total concentration of protein and ligand; $\Delta\delta_{obs}$ is the change observed in the shift from the reference spectrum, while $\Delta\delta_{max}$ is the maximum shift change obtained upon saturation and is generally calculated from the fitting, since it is not always possible to measure it experimentally.

This method is applicable only in the case of low affinities (high µM or mM range) in case of fast exchange regime (K_{ex}>>Δω) [118].

HSQC has a high sensitivity which allows detection of very low affinities of binding (mM range). The main limitation in the usage of protein-observed NMR in structural biology is due to the size of the protein and the large amount of protein required (~ 0.1 mM range). High-quality HSQC spectra can be obtained only for proteins with MW < ~40 KDa, otherwise peak overlapping and weakening of signal is usually experienced. In some cases, it is possible to circumvent these limitations with predeuteration of the protein or using TROSY experiments for larger molecules [120].

1.4.1.2 Ligand observed NMR

In the ligand observed NMR, the observable signals are those from the ligand [121]. It is a more cost effective technique since there is no need to label the protein and small amount of ligand and protein is required for the analysis. It is widely applied in fragment screening to select and validated low affinity binders [122]. In contrast to the protein observed NMR, ligand-observed methods do not have an upper limit for the MW of the protein analysed; instead, there are limitations when applied to protein with MW < 10 KDa since with a small protein the changes on the parameters monitored by NMR on the small molecule would be smaller, therefore, the response detected may be too small to be detected even in a case of binding event [123].

In this thesis three different techniques were used: saturation transfer difference (STD) [124], Water-Ligand Observed via Gradient SpectroscopY (WaterLOGSY) [125] and relaxation-edited Carr-Purcell-Meiboom-Gill (CPMG) [126].

In STD experiments, methyl groups of the protein, generally buried in the core of the protein and often close to potential binding pockets, are saturated by a selective pulse. If a ligand is added to the protein and is interacting with it, magnetization of the ^{1}H nuclei is transferred to the adjacent nuclei through dipole-dipole interactions (also known as Nuclear Overhauser Effect, or NOE) and to the ligand located in proximity. If the ligand is binding to the protein, its signals during this ‘on resonance’ step are fully saturated

while non-binding compound will not be affected. Then an 'off-resonance' spectrum is recorded with pulses where no saturation is performed. Subtraction of the 'on-resonance' spectrum from the off-resonance spectrum will generate the STD spectrum with signal only from the saturated ligand and protein.

In WaterLOGSY experiments, the same physical principles underpinning the STD experiment apply, however in this case it is bulk water that is excited, rather than protein methyl groups. Magnetization can then be transferred from water to the bound ligand via NOE effect, either directly or via solvent-exchangeable protons on the protein. If the ligand is only interacting with bulky waters it will experience a low NOE of positive sign due to its much faster tumbling. Small molecules bound to the protein will experience lower tumbling and this will generate a stronger NOE of negative sign. For convenience, signals are generally phased in order to have a negative signal for the free ligand, meaning that bound ligands will exhibit positive or less negative signals compared to a control sample of ligand in the absence of protein.

CPMG experiments rely on the observations that the signal obtained upon addition of macromolecules to the ligand sample is reduced if the macromolecule is interacting with the ligand. The principle of this phenomenon is in the different relaxation times experienced from small molecules and macromolecules. Macromolecules show faster dynamics in solution and as a result relax in a shorter time (T_2) than a small molecule. Therefore, if the ligand binds to the macromolecule, the signal of the ligand population that is bound to the small molecule is broadened, filtered out, and ultimately reduced in intensity. The technique is dependent on the spin-lock sequence applied (typically 60 ms) which requires to be comparable or longer to the T_2 of the bound ligand.

1.4.2 X-ray crystallography

X-ray crystallography is a powerful technique to gain structural insights into the molecular basis of the protein-protein and protein-ligand interactions [127]. The bottleneck of the process is in the possibility to crystallize the protein or the protein complex. Different techniques and strategies have been developed in recent years in

order to improve crystallization of proteins but despite this, many proteins are still recalcitrant to the process. On the other side, advances in computational methods and robotics applied to crystallography have seen a notable increase in throughput, bringing this technique to be routinely used in structural biology and drug discovery [128]. The application of crystallography in drug discovery relies on the possibilities to obtain ligand bound crystals of the protein of interest, through either co-crystallization or soaking [129]. Structure determination can be assessed from the diffraction pattern generated by the crystals upon X-ray irradiation. Processing of the diffraction data collected leads to an electron density map reconstruction. Interpretation of this map determines the protein structure. Observation of differences in electron density between the holo and *apo* form of the protein can bring to the identification of an extra electron density for the bound ligand.

1.4.3 Bio-layer interferometry (BLI)

BioLayer interferometry (BLI) is a label-free optical technique used to study in real time the binding of biomolecules [130]. It measures interference pattern of white light in response to the interaction between molecules. One interaction partner is immobilized (for example, using affinity interactions) on a biolayer at the top of a biosensor tip and can potentially interact with the analyte in solution. White light is pulsed across the biosensor till the biolayer and is then reflected back. Light reflected back is analysed by a CCD array detector. If the thickness of the biolayer at the top of the tip changes as consequence of interaction between ligand and analyte, the pathway of the light reflected increases and this will result in a wavelength shift of the interference pattern of the reflected light. The pattern of reflected light is therefore a function of the thickness of the bio-layer at the top of the tip. Detection of the wavelength changes over time generates the sensorgram.

1.4.4 Surface plasmon resonance (SPR)

Surface plasmon resonance (SPR) is a label-free optical technique used to monitor binding between macromolecules and small molecules or other macromolecules [131].

SPR is a phenomenon that occurs when polarized light incident on a thin metal layer has all the photons with an equal momentum of the surface plasmon of the metal layer. This phenomenon verifies at a specific angle, called resonance angle θ, and the optical energy is coupled into the evanescent waves of the metal plasmon which irradiate for ~100-200 nm out of the surface of reflection. Consequently, the intensity of the reflected light is reduced. Any change at the interface between metal and solution will change the momentum of the surface plasmon and, therefore, the resonance angle. Thus, in SPR the observable is the resonance angle of the light reflected from a metal surface of a prism. The SPR response can shift as result of interactions happening at the metal/solution interface. In SPR, a microchip coated with gold (or any other relevant metal) is generally used and it provides the platform for immobilization of the analyte.

Immobilization of the protein in BLI and SPR can be performed through covalent coupling with the chip surface using free amine, thiol or carboxylic group. The issue related to this chemical method is the resulting random orientation of the protein. The usage of affinity tags (as His-tag) can be a valid alternative to overcome the orientation problem. Also biotinylation can be used to immobilize the protein on a streptavidine coated chip, relying on the strong affinity between biotin and streptavidine. Biotinylation is not site specific, this approach may suffer too of the orientation problem of the binding site.

BLI and SPR can be both used for K_D estimation either via analysis of kinetic parameters or steady state analysis. In the first case, suitable for stronger affinity, K_D can be calculated as the ratio between the kinetic on-rate and off-rate constants. In the second case (preferred in case of weak interactions), K_D can be extrapolated from the fitting of the equilibrium responses at different concentrations of ligand.

1.4.5 Isothermal titration calorimetry (ITC)

Isothermal titration calorimetry is a biophysical technique that allows the measurement of the heat exchanged during a binding reaction between two components (for example, protein and small molecule or peptide) [123]. The instrument is made of two cells: a

reference cell and a sample cell, both inserted in an adiabatic jacket to minimize the exchange with the environment. The temperature between the two cells needs to be kept constant during the experiment and the power supplied to maintain this balance can be considered as a measurement of the heat, released or absorbed, during the binding reaction. ITC equipment is armed with a syringe that allows iteratively injections of the ligand into the sample cell, in order to measure the heat of binding. The area of the peak associated to each injection is proportional to the heat released (exothermic) or absorbed (endothermic) during the reaction. The measured heat upon each injection is normalized per mole of injectant (typically the ligand) and plotted *vs.* the molar ratio [L]/[P] to generate an isotherm. Analysis of the isotherm allows the extrapolation of important thermodynamic parameters: the association binding constant (K_a, which is obtained from the slope of the isotherm at the inflection point), ΔH of binding (obtained from the extrapolation of the binding isotherm at [L]/[P] = 0) and the stoichiometry n value (the molar ratio at the inflection point). ΔG and ΔS can be calculated considering equations (4) and (5):

$$\Delta G = -\mathrm{RT}\ln K \qquad (4) \quad \text{and} \quad \Delta G = \Delta\mathrm{H} - \mathrm{T}\Delta\mathrm{S} \qquad (5)$$

1.4.6 Differential scanning fluorimetry (DSF)

Differential scanning fluorimetry (DSF) [131], also known as thermal shift assay, is a technique that monitors the protein denaturation process upon increase of temperature [132]. DSF allows monitoring the temperature-dependent unfolding process of a protein using an environmentally-sensitive fluorescent dye, generally the SYPRO Orange. A fluorophore on the dye absorbs light at a definite wavelength and re-emits radiation at a longer wavelength (fluorescence). Fluorescence is quenched in water, but it increases when the dye binds to hydrophobic surfaces which get exposed as the protein unfolds at increasing temperature. Therefore, monitoring the fluorescence signal from the dye, it is possible to follow the denaturation process. Relative fluorescence unit are plotted against the temperature and a sigmoidal-type curve is generated. The inflection point of this curve determines the Tm, that can be extrapolated using the Boltzmann equation [132].

This technique is largely used to test the stability of a protein under different conditions, which provides helpful information to aid protein crystallization. In addition, DSF is also used in early lead discovery as primary fragment screening technique. Indeed, when a compound binds a protein, it generally stabilizes the protein, with consequent positive shifts on the Tm of the denaturation process. It has however been found that also destabilizing compounds can be assessed as true binders [133 134].

1.4.7 AlphaLISa

AlphaLisa is a biochemical assay that relies on the bead-based technology developed by PerkinElmer [135]. Two kinds of beads are typically used in this assay: donor and acceptor beads. Donor beads contains a photosensitizer (phthalocyanine) which upon light irradiation at 680 nm is responsible for the excitation of a ground state O_2 into singlet oxygen. Singlet oxygen is long lived but, upon interaction with the solvent, its lifetime is reduced to 4 µs, and as a result it can then irradiate within a distance of ~200 nm. If the acceptor bead, which contains europium chelates, is located in proximity of the singlet oxygen, it will be excited from the energy released in the decay of the excited state of O_2, generating subsequent emission at a specific wavelength (615 nm). The distance between acceptor and donor beads will depend from the interaction between the molecules that are bound to the acceptor and donor beads. In the assay developed in this thesis the donor beads are covered with streptavidin that binds biotinylated protein, while the acceptor beads are covered with anti-FLAG antibody that binds the Flag-tagged histone peptide. The emission spectrum for the AlphaLisa assay is in a narrow defined range, which reduces interference from the bulk solution. The illumination (680 nm) and irradiation (615 nm) wavelengths are very specific so interference from the background are minimized and a good signal-to-noise is warranted. AlphaLISA is a high-throughput assay which allows optimal versatility in the design. Its high sensitivity and multiplicity of derivatization sites on the beads facilitate detecting even low affinity interactions at very low concentrations of ligands, as these can be artificially enhanced by proximity induced multivalency of the interaction on the beads.

1.4.8 MODified™ Histone Peptide Array

The Modified histone peptide array is a tool used to screen proteins and their interactions with histone peptides carrying multiple post-translational modifications. The array is constituted of a coated slide glass where 19-mer peptides conjugated to cellulose are spotted onto. Each array contains a double copy of 384 unique histone modifications of H2A, H2B, H3 and H4. Each peptide can contain up to four different combinations of PTMs in order to test influence of adjacent modifications on the recognition of a specific PTM. The high peptide density ensures detection of interaction also at low binding affinities. The array works much like a Western blot. A protein of interest is first incubated with the array for a short time, typically two hours. Subsequently, incubation with a primary antibody is carried out to bind the interacting protein. Incubation with a conjugated secondary antibody (horseradish peroxidase, HRP) is performed and then chemiluminescent detection is performed using a X-ray film. A schematic representation of the process is illustrated in Figure 1.13.

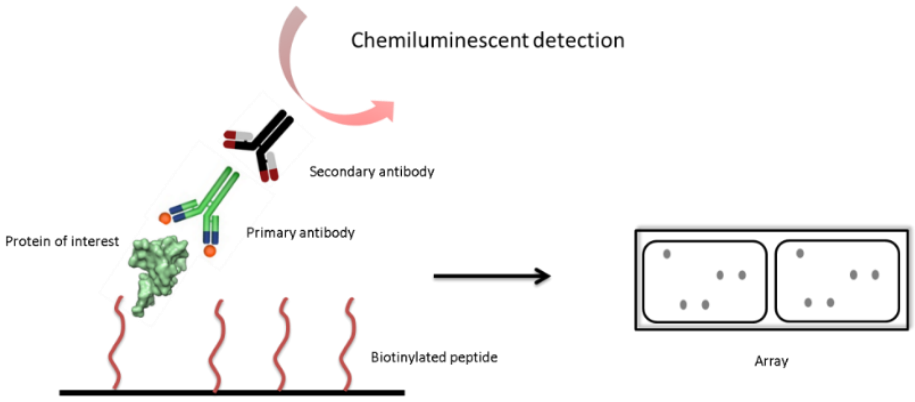


Figure 1.13. Simplified representation of the histone peptide detection on array.

1.5 AIMS OF THE PROJECT

The aim of this project was to assess the ligandability of a class of reader domains, poorly investigated up to date. As model of study were chosen the PHD fingers of BAZ2A and BAZ2B. To achieve the goal two main aims were pursued.

The first aim was to shed light on the recognition mode of the PHD zinc finger of BAZ2A and BAZ2B toward the H3 histone-binding partner. The work performed to achieve this aim is described in Chapter 1, and it was performed in collaboration with Dr. Alessio Bortoluzzi and Dr. Xavier Lucas (Ciulli Group). The paucity of information on the histone recognition mode of these PHD domains inspired the work presented in the first result chapter of this thesis.

In the second part of the project, the research focussed on interrogating the ligandability of the PHD fingers of BAZ2A and BAZ2B with small molecule fragments. Indeed, from literature, bromodomains of BAZ2A and BAZ2B had been perceived as ligandable [40, 136-139] but no study have yet addressed the ligandability of the PHD zinc finger domains of these proteins. It was decided to perform a fragment-based approach to address this question, with the objective to identify useful targeting points for chemical probe design in future.

In a third part of the project, a study was performed to begin to elucidate the interaction of the PHD-BrD of other two proteins containing a PHD-Brd reader binding module, BAZ1B and TRIM66, towards their potential histone binding partner. Both these proteins are known to be involved in diseases but their biological role is not been completely elucidated [109, 115, 116, 140]. Herein, we believe that gaining insights on the structural basis on the binding of their reader domains with the cognate binding partner would help towards the identification of potential chemical probes able to target their histone interactions.

Ultimately, chemical probes could be advantageous in disentangling the complex biological networks in which these proteins are involved and could provide key knowledge for the development of future modulator of the interaction to be used as potential molecular therapeutics.

CHAPTER 2.
Molecular basis of histone tail recognition of PHD of BAZ2A/B

2.1 MOTIVATION OF THE WORK

Understanding the structural basis of the interaction of a protein with its natural binding partner can lead to the identification of hotspots of interaction, involved in the molecular recognition. This knowledge could also steer the discovery of potential modulators of the protein-protein interaction (PPI). Indeed, for those PPI where the binding involves linear peptidic epitope, as in the case of PTMs recognition, the hotspots of interaction have the potential to be used to build a rational design for the development of chemical probes able to target this PPI [87].

Drug discovery in the field of epigenetic readers and reader domains has largely benefited from this structure-based approach that has driven successful results in the identification of small molecules able to disrupt protein-protein interactions. Remarkable examples are the generation of potent inhibitors of bromodomains [80] or of the histone N-methyl-lysine demethylases (KDMs) [141]. For instance, the small molecule JQ1(+) deeply binds to a pocket on the bromodomains of BET proteins (BRD2, BRD3, BRD4 and BRDT) that recognizes specifically the side chain of acetyl-lysine residues [142]. In this case, this pocket could be considered the hot-spot of the interaction.

While successful results have been achieved targeting bromodomain readers, less progresses have been reported for targeting other reader domains. For example, PHD zinc fingers, often found in proximity of a bromodomain, amongst others, constitute a class of readers poorly addressed from the drug discovery point. To date, there are only two cases, in literature, assessing a systematic study of their ligandability that led to the identification of small molecules able to disrupt histone PPI by acting as competitors with the histone binding pocket [56, 57]. For these readers, the majority of the efforts has been placed in exploring and understanding the readout of the PTMs of the histone tails

by PHD zinc fingers.

The work presented in this chapter aimed to shed light on the structural basis of the molecular recognition of the H3 histone tail peptide by the PHD zinc finger of BAZ2A and BAZ2B, chosen as model systems for this type of study. The ultimate goal of the research was to furnish structural knowledge to guide a drug discovery campaign towards the identification of chemical probes able to disrupt the reader-histone interactions of this class of proteins.

This study was performed in collaboration with two postdocs of the Ciulli group, Dr. Alessio Bortoluzzi and Dr. Xavier Lucas, who contributed toward the preset goals.

2.2 CRYSTAL STRUCTURE OF PHD OF BAZ2A IN COMPLEX WITH H3 10-MER PEPTIDE

Previous work in the lab had led to the identification of the unmodified H3 histone tail as favourite binding partner for the PHD zinc finger of the two proteins BAZ2A/B [45]. Single methylation on H3K4 was only partially tolerated, whereas double and triple methylation abrogated binding [45]. Other modifications such as phosphorylation on T3 or methylation on R2 of the H3 histone led to a decrease in binding affinity [45]. It was also observed that different lengths of the H3 peptide, resembling the N-terminal tail recognized by these PHD domains, led to different binding affinities. Isothermal titration calorimetric (ITC) experiments reported ~4-fold higher affinity for both proteins BAZ2A/B PHD towards H3 10-mer (BAZ2A/B K_D ~ 50 µM) compared to the H3 5-mer peptide (BAZ2A/B K_D ~ 200 µM) [46]. HSQC NMR experiments performed by Dr. Alessio Bortoluzzi on both proteins BAZ2A/B PHD in complex with H3 10-mer and H3 5-mer showed equivalent CSPs maps for the two peptides. No extra shifts were detected for the longer peptide [46]. HSQC NMR experiments also highlighted that the binding mode was essentially conserved between the two proteins, as expected, since BAZ2A/B PHDs harbor 66% of sequence identity.

To investigate the interaction of the PHD BAZ2A/B with their cognate histone binding partner H3 at atomic detail, I synthesized and purified a 10-mer peptide (ARTKQTARKS),

that resembles the N-terminal tail of the histone protein H3, to investigate its binding mode at atomic detail.

The *apo* form structures of the PHD of BAZ2A/B were previously described [45]. It was observed that in these crystal forms the histone binding pocket for the PHD of BAZ2A was solvent-exposed and free from crystal contacts, at least in two of the four protomers of the asymmetric unit (ASU). In contrast, in the BAZ2B PHD crystal packing, access to the pocket was partially occluded by crystal contacts made with protomers from the adjacent ASU.

In light of these observations, to elucidate the molecular details of histone H3 N-terminal tail recognition, I soaked H3 10-mer peptide into the apo form crystals of the PHD of BAZ2A.

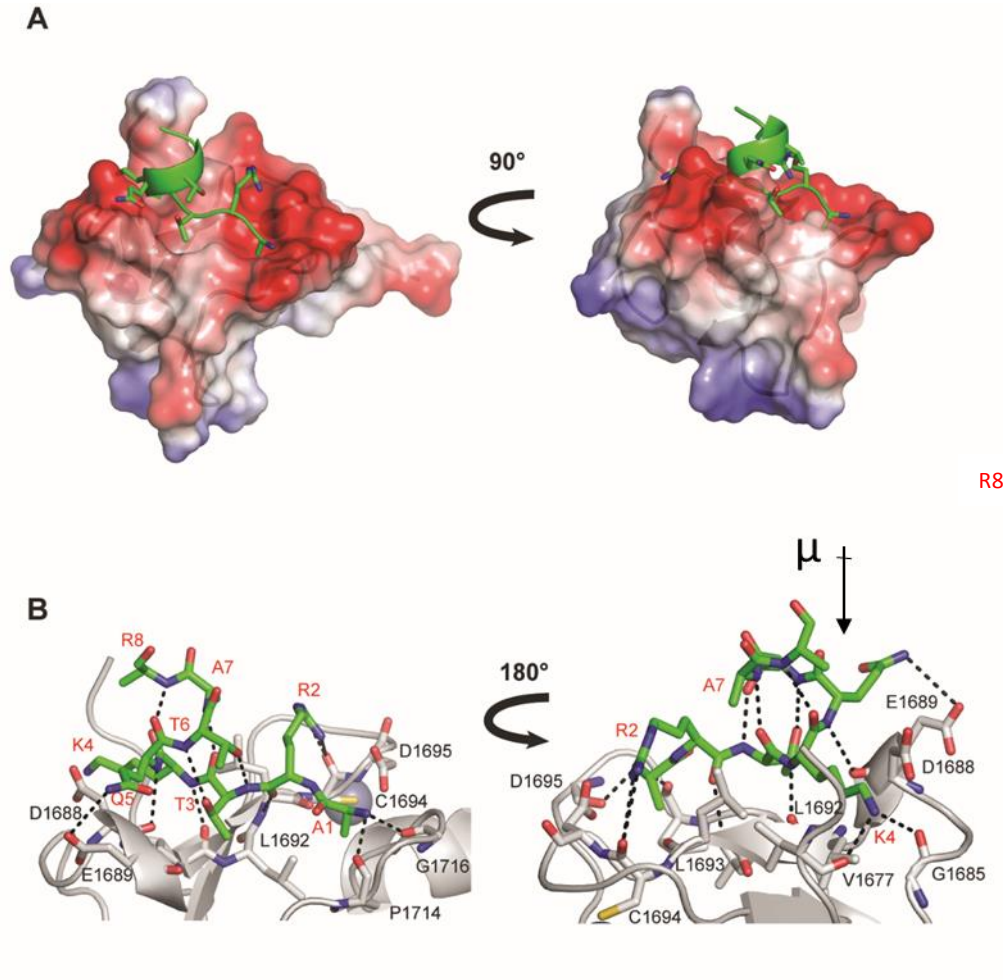
Apo form crystals of BAZ2A PHD (figure 2.1) were obtained mixing equal volume of protein and crystallization buffer on a sitting drop plate. Protein was concentrated up to 6 mg/mL in 20 mM Tris pH 8.0, 200 mM NaCl, 5 mM DTT and 20 µM $ZnCl_2$. Crystallization buffer was 2.2 M Na/K phosphate buffer pH 8.5. Crystals were left to grow for one week at 20 °C, till they reached their final size. They were subsequently harvested from the mother solution and soaked in 10 µL solution of reservoir containing 2 mM of H3 10-mer peptide. After ~16 h, crystals were flash-frozen in liquid nitrogen after few seconds of soaking in a solution containing 1.6 mM H3 10-mer peptide and 20% glycerol for cryoprotection.

Figure 2.1. Crystals of BAZ2A PHD

Drop containing single crystal of PHD BAZ2A in 2.2 M Na/K phosphate buffer pH 8.5.

The structure was solved by isomorphous refinement with the *apo* form of BAZ2A PHD where the search model for the calculation was the chain A (PDB: 4QF2) [45]. Details are reported in chapter 6.5.

Model validation was carried out using MolProbity [143]: each protomer had excellent stereochemistry confirmed by no outliers in the Ramachandran plot. Crystallographic R-factors for the overall structure were 18.5% and 23.5%, respectively, for R_{work} and R_{free}. Table 2.1 summarizes all the crystallographic data processing and refinement statistics for the crystal structure of BAZ2A PHD in complex with H3 10-mer peptide.



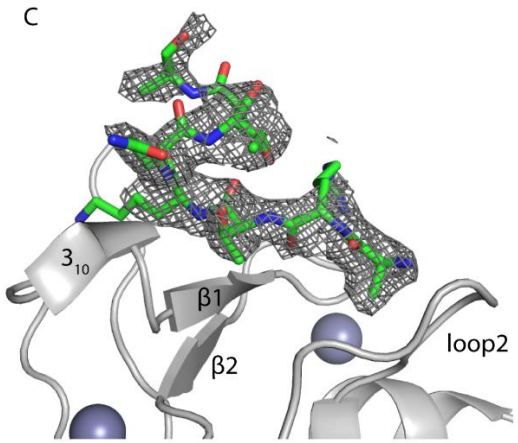


Figure 2.2. Crystal structure of PHD of BAZ2A in complex with H3 10-mer [46].

(**A**) BAZ2A PHD (surface representation) in complex with H3 10-mer (green ribbon). Protein surface is coloured according to the electrostatic potential: blue for positive region and red for negative regions. (**B**) Close-up view of the interactions between BAZ2A PHD (grey cartoon) and the H3 10-mer peptide (green sticks, labelled in red). Residues of BAZ2A PHD interacting with the peptide are highlighted as sticks and labelled in black. The arrow indicates the direction of the macrodipole μ associated with the helical formation. (**C**) The F_o–F_c map for H3 10-mer is shown in grey and contoured at 2.5σ.

Analysis of the structure led to the identification of the main spots involved in the binding. The peptide residues A1-K4 form hydrogen bonds with the first β-strand of the protein BAZ2A PHD, anchoring the peptide into the pocket. In detail, the first residue of H3, A1 is involved in hydrogen bonds with the carbonyl group of the P1714 and G1716, while the rest of the peptide is stabilized through hydrogens bonds with the backbone of the residues L1693, L1692, L1691, D1688 and E1689 (Fig. 2.2 B). The methyl groups of A1 and T3 are involved in hydrophobic contacts with the protein as well as the side chain of K4 that is stabilized by hydrophobic interactions with L1691 and V1677 plus hydrogen bonds with the carbonyl group of the backbone of G1685 and V1677 and, additionally, by long-term interaction with the carboxylic group D1688. The side chain of R2 is involved in electrostatic interactions with the side chain of D1695 and in hydrogen bond with a structural water, located between the carbonyl group of L1693 and the carboxylic group of D1695; this water also mediates important interactions of the protein with the backbone amide NH of R2 (Fig. 2.3.). The peptide adopts a helical fold after residue K4 and till R8, forming an α-helical loop stabilized by intramolecular hydrogen bonds between the carboxylic groups of T3 with the amide group of A7 and

between K4 and the amide group of R8 (Fig. 2.2. A-B). Also the hydroxyl group of the side chains of the residues T3 and T6 are contributing to the stabilization of the peptide through the formation of intermolecular hydrogen bonds, respectively, with the amide group of T3 and T6 (Fig. 2.2 B). The helical conformation is associated with the formation of a helix macrodipole which is the sum of the dipoles directed from the negative pole of the oxygen of the peptidic bond to the positive pole of the hydrogen. The arrow in figure 2.2 B illustrates the direction of the dipole. There is no interpretable electron density for the side chain of R8 after Cβ, as well as for the residues K9 and S10, implying that these are disordered residues (Fig. 2.2 C). Inspection of the asymmetric unit found the peptide bound to two of the four protomers, chains A and C, exactly as expected. The histone pockets of these two protomers are free from any crystal contact that may interfere with the structural rearrangement of the peptide itself. Therefore, it is highly unlikely that the helical fold is an artefact of the crystal packing and soaking procedure. Conversely, crystal packing occludes the binding sites of the other two protomers, chains B and C, where no peptide is found bound.

No conformational changes were observed between the bound and unbound form of the PHD of BAZ2A. This was expected since the PHD pockets, as in other methyllysine readers, are defined as static pockets which induce conformational changes into the histone peptide without significant perturbation on their own structure [8].

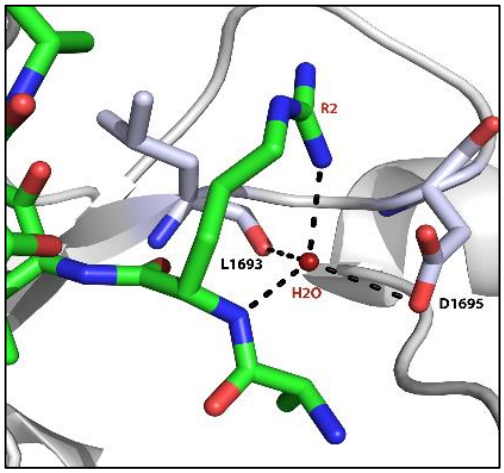


Figure 2.3. Detailed view of H3R2 interacting with BAZ2A PHD.

Close-up view of the intermolecular interactions mediated by the structural water (red sphere) between R2 of H3 10-mer peptide (green) and L1693 and D1695 of BAZ2A PHD (grey).

This structure (PDB: 5T8R) was compared with the structure, previously solved in the Ciulli lab [45], of the PHD of BAZ2A in complex with H3 5-mer peptide ARTKQ (PDB: 4Q6F) (figure 2.4).

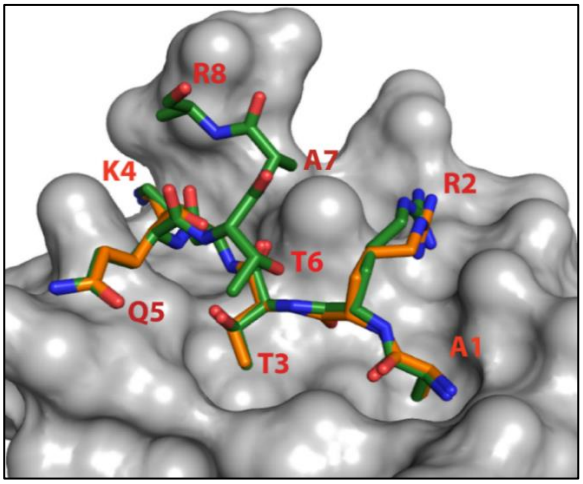


Figure 2.4. Superposition of PHD BAZ2A in complex with H3 10-mer and 5-mer peptides.

PHD of BAZ2A (surface in grey) in complex with H3 5-mer peptide (orange, 4Q6F) and H3 10-mer peptide (green, 5T8R). Peptide residues are labelled in red.

Superposition of the structures (Fig. 2.4) did not highlight any further interaction established by the longer H3 10-mer peptide directly with the protein. Indeed, residues T6-S10 were not involved in binding except for the methyl group on A7 that forms a long range hydrophobic contact with the side chain of L1693 (Fig. 2.2 B) [46].

In light of these results, a reasonable hypothesis formulated to justify the increase of affinity of the longer H3 peptide was the stabilization of the helicity of the peptide within the histone pocket. This helicity is required to avoid clashes of the peptide onto the protein [46].

2.1.1 Co-crystallization efforts of the PHD of BAZ2B -H3 10-mer complex

In order to validate our observations, arising from NMR data, on the binding mode of the H3 N-terminal tail by the PHD of BAZ2B as well, I performed several crystallization trials. As mentioned above, the inspection of the histone pocket within the apo form of the crystals of the PHD of BAZ2B (PDB: 4QF3) showed this pocket as partially occluded

by crystal contacts with protomers of the symmetry related unit. This discouraged soaking attempts, and I preferentially undertook co-crystallization.

Three sparse matrix screening were used: Proplex, Classic and PegRX, each at 4°C and 20°C.

Two different protein concentrations were used ~6mg/mL and ~14 mg/mL in a protein: peptide ratio of 1:5. At 20°C the majority of the drops showed amorphous or microcrystalline precipitate. Instead, at 6mg/mL drops appeared mainly to be clear as well as at 4°C, despite the high concentration of protein.

Interesting hits were obtained from one of the conditions (G1 from classic screening: 0.1 M HEPES sodium salt pH 7.5, 2% PEG400 and 2M $(NH_4)_2SO_4$) (Fig. 2.5). The high concentration of the salt raised some doubts about the nature of the crystal hits. This was elucidated using the Izit crystal dye (Hampton) (Fig. 2.5 B). Izit is a small molecule dye which fills the solvent channels in protein crystals, colouring the crystals in blue. If the crystals are made of protein molecules, this dye would leave a clear drop with blue crystals. Salt crystals do not possess these large solvent channels and the dye cannot enter the crystal. Thus, in case of a salt crystal this will remain clear and the drop will appear blue. Izit crystal dye was added to the crystal found and after ~1 h incubation the crystals remained clear (Figure 2.5 B) confirming that these were salt crystals.

Figure 2.5. Izit test with crystal hits.

(**A**) Crystal hits observed from G1 condition in Classic screening at 20°C. (**B)** Crystal hits after Izit uniformly coloured the drop.

The majority of the drops from these co-crystallization screening showed only microcrystalline formations, heavy precipitation or clear drops. No other potential hits were observed.

2.3 CHANGE IN H3 N-TERMINAL TAIL HELICITY CORRELATE WITH DIFFERENT BINDING AFFINITIES FOR BAZ2A/B PHD ZINC FINGER

To gain a better understanding of the histone molecular recognition by the PHD of BAZ2A, the energetic contribution of the residues on H3 involved in binding was investigated [7]. An alanine scanning was performed, by Manuel Blank and Dr. Alessio Bortoluzzi in the Ciulli lab, mutating to alanine residues from 2 to 6 on the H3 peptide [46]. This experiment highlighted that that residues R2 and T3 were essential for binding [46] while K4-T6 were found not critical for the binding but, unexpectedly, double mutation to alanine (H3 - ARTAATARKS) increased the affinities for both BAZ2A/B PHDs compared to the wild-type H3 [46]. These results were unexpected since the removal of the K4 sidechain was predicted to lead to a decrease of binding as consequence of loss of contacts between peptide and PHD (Fig. 2.2 B).

Therefore, the hypothesis was that the K4A and Q5A mutations could stabilize the helicity of the peptide, indeed, alanine is known to be a helix-stabilizing residue [144]. To support this idea, K4G and Q5G double mutations on H3 (ARTGGTARKS) showed remarkable reduction of affinity in the ITC experiments performed by Dr. Alessio Bortoluzzi [46]. This was expected since glycine has the lowest helix propensity among all the natural amino acids [144].

To further validate these observations and gaining more details at molecular level of this interaction, I solved the crystal structure of the PHD of BAZ2A in complex with H3 10-mer AA mutant at 1.9 Å.

The apo form crystal of BAZ2A PHD was soaked overnight in a solution containing crystallization buffer (2.2 M Na/K phosphate pH 8.5) and 2.5 mM H3 AA mutant 10-mer peptide and subsequently flash frozen in liquid nitrogen after few seconds soaking in the reservoir with 20% glycerol as cryoprotectant. Structure of the complex was solved (figure 2.6) following the same procedure described in chapter 6.5, as for the complex BAZ2A PHD-H3 10-mer WT. Crystallographic data of the AA mutant are reported in Table 2.1.

Inspection of the structure led to the consideration that the helical binding mode for H3 10-mer peptide AA mutant was conserved compared to the wild-type (figure 2.7 A-C). Superposition of the two peptide bound forms of BAZ2A PHD showed that no changes

in the conformation of the PHD pocket were observed between the proteins, which perfectly superimposed. The hydrogen bond distances between protein and peptide were analysed and for the first three residues (ART) they were the same in both peptides (values are reported in figure 2.7 B). Remarkable differences in H-bonds distance were observed between protein and peptide in the region of the mutated residues (figure 2.7 B and C). Indeed the intermolecular H-bond involving the N-H of H3K4 and H3A4 are respectively of 3.0 Å and 2.4 Å and the H-bonds between the C=O in D1688 and NH from H3Q5 and H3A5 are respectively 3.4 Å and 2.5 Å. Shorter distances were observed also for the intermolecular peptide H-bonds within the double mutant peptide compared to the wild-type (figure 2.7 A). These shorter distances suggested a tighter conformation of the peptide within the pocket which could account for the increased affinity for the alanine mutant peptide (H3 AA mut).

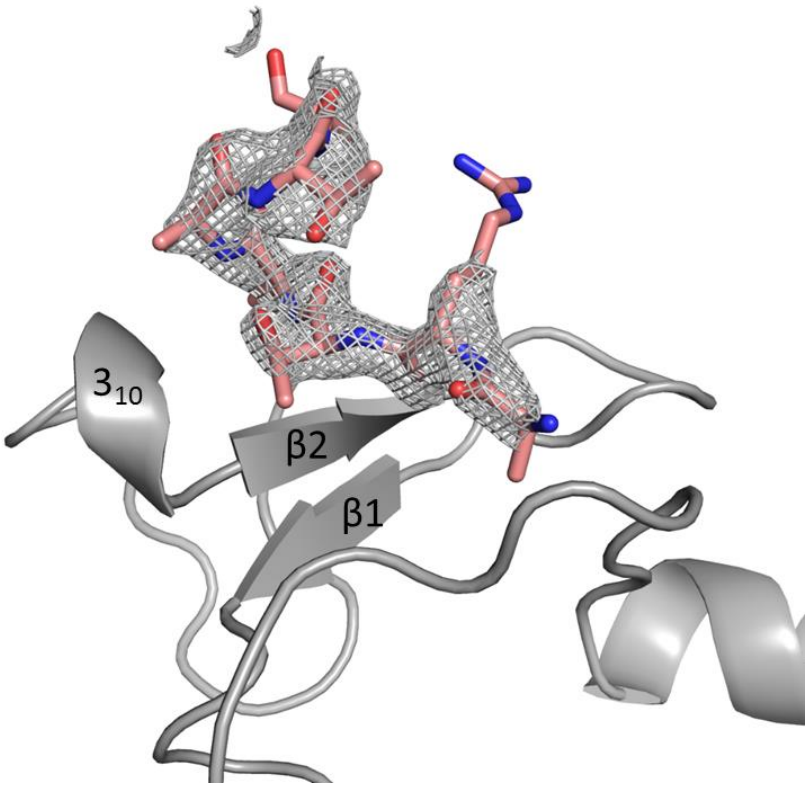


Figure 2.6. Crystal structure of BAZ2A PHD in complex with H3_AA_mutant.

Crystal structure of BAZ2A PHD (in gray) in complex with H3 10-mer double mutant (in pink). (***A***) The F_o–F_c map contoured at 2.5σ (shown in gray) for H3 10-mer AA mutant.

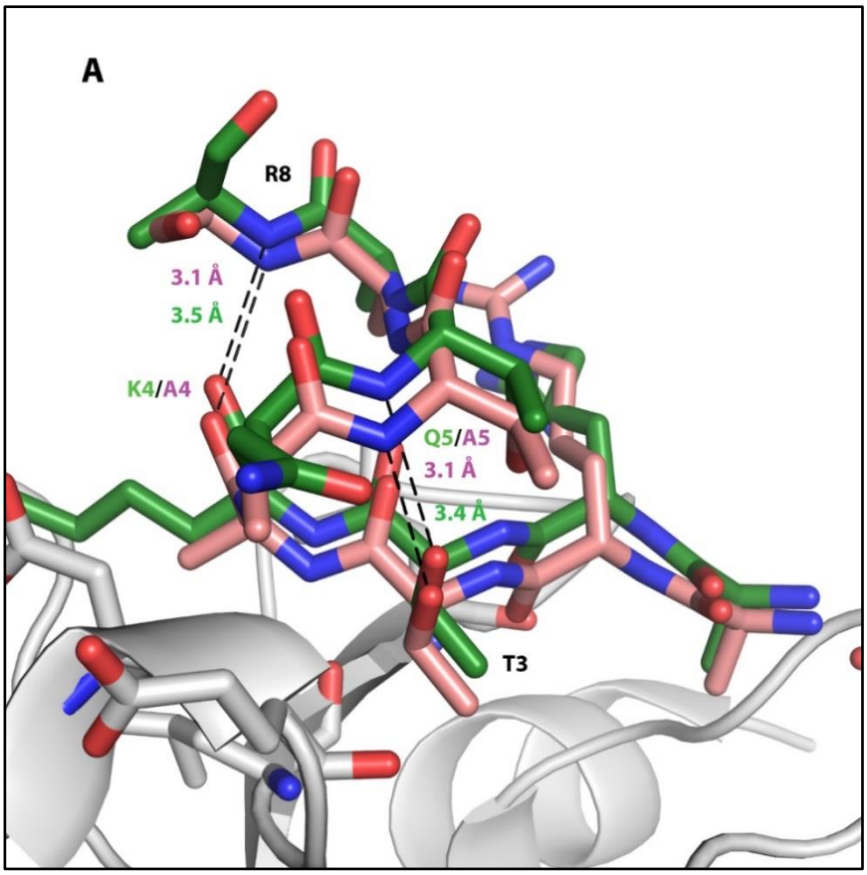


Figure 2.7. A. Superposition of the complex of BAZ2A PHD with H3 10-mer wild-type and mutant peptide.

Superposition of PHD BAZ2A (grey) in complex with H3 10-mer (green sticks, 5T8R) and H3 10-mer AA mutant (pink sticks, structure not deposited). Intramolecular peptide bonds are drawn in dash and distances are reported in Å, in green for H3wt and in pink for H3 mutant.

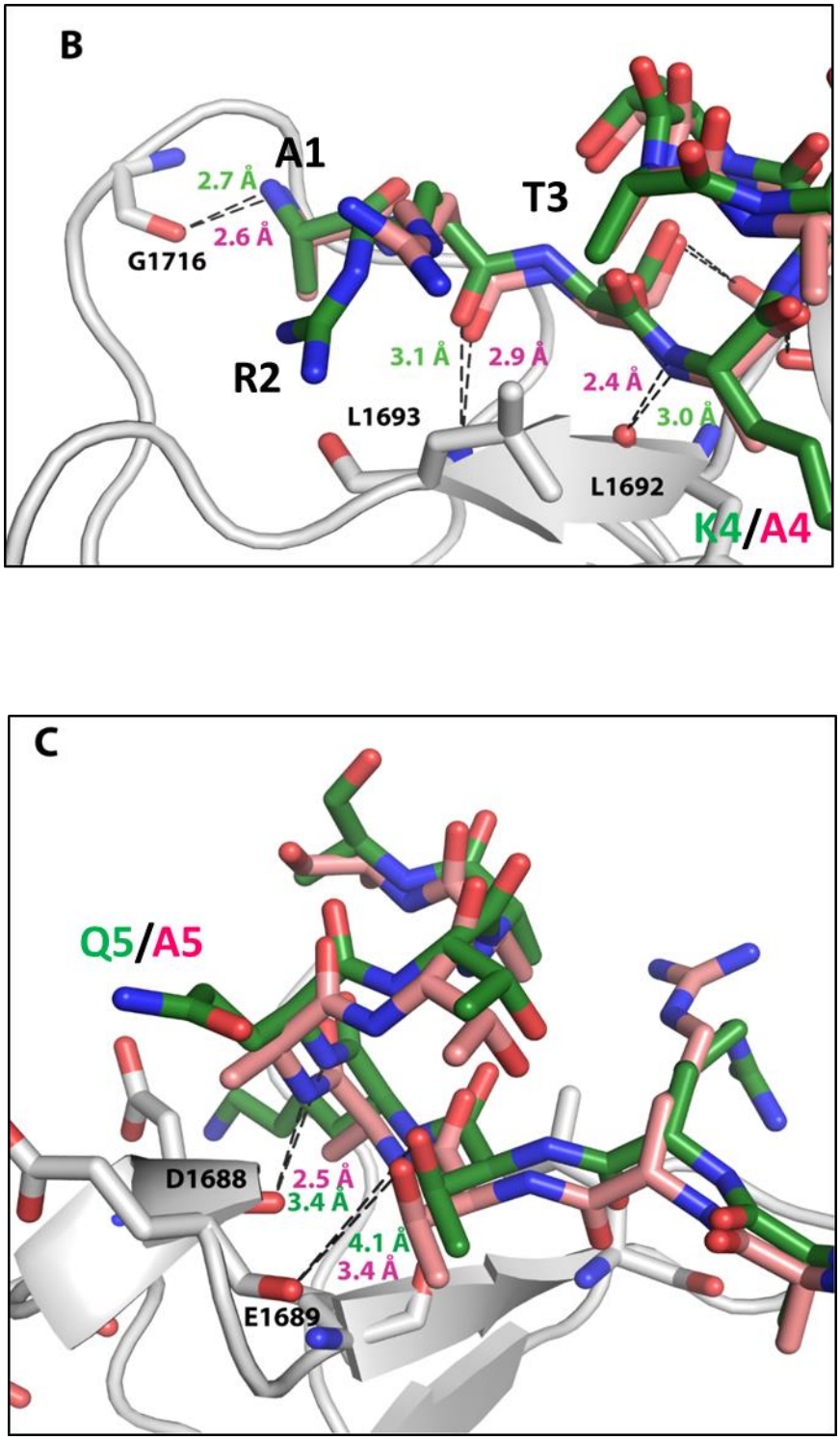


Figure 2.7 B-C. Superposition of the complex of BAZ2A PHD with H3 10-mer wild-type and mutant peptide.

Superposition of PHD BAZ2A (grey) in complex with H3 10-mer (green sticks, 5T8R) and H3 10-mer AA mutant (pink sticks, structure not yet deposited). Hydrogen bonds are in dashed lines and distance in Å. In panels B and C, differences of the intermolecular distances. Residues are labelled in black and the peptide is in green for the wt H3 10-mer and in pink for the H3 10-mer AA mutant.

Protein ID	**BAZ2A PHD**	**BAZ2A PHD**
Ligand	ARTKQTARKS	ARTAATARKS
Processing statistics		
Space Group	$P4_32_12$	$P4_32_12$
Unit cell parameters		
a, b, c (Å)	72.61, 72.61, 99.43	72.7, 72.7, 99.9
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution limits (Å)	45.62 -2.4 (2.49 – 2.4)	45.7 -1.9 (1.9 – 1.94)
Unique observations	10901 (1132)	21777 (1371)
Completeness (%)	99.6 (100)	99.9 (99.8)
Redundancy	5.4 (5.8)	9.7 (10.3)
R_{merge} (%)	10.2 (72.8)	9.8 (96)
I/σI	9.3 (2.3)	12.3 (2.6)
$CC_{1/2}$ (%)	99.4 (71.8)	99.5 (79.4)
Wavelength (Å)	0.9762	0.9762
Refinement statistics		
R_{work}/R_{free} (%)	18.5 / 23.5	19.3 / 24.5
Average B factors $(Å)^2$	52.13	52.13
RMSD bond (Å)	0.01	0.02
RMSD angle (°)	1.50	2.50
Ramachandran Statistics		
Favored (%)	97.2	94.9
Allowed (%)	2.8	9
Outliers (%)	0.0	0.9
PDB	5T8R	XXXX

Table 2.1. Crystallographic data processing and refinement statistics [46].

Value in parenthesis are relative to the highest resolution shell.

2.4 IDENTIFICATION OF THE ACIDIC WALL ON THE PHD OF BAZ2A/B AND ITS ROLE IN THE MOLECULAR RECOGNITION OF H3 PEPTIDE

The helical conformation adopted by the peptide was found as a critical feature essential to avoid the clash of the histone peptide with the 3_{10} helix of the histone pocket. Moreover, structural analysis led to the observation that the 3_{10} helix or a small turn, preceding the first β-strand of the PHD, is characteristic of all those PHDs recognizing the histone tail in α-helical conformation. Bioinformatic analysis performed by Dr. Bortoluzzi, highlighted that this region is often characterized by an acidic patch which was found mutually exclusive with the conserved tryptophan, needed for the K4me recognition. This region was named as 'acidic wall', since it was found to contain at least one acidic residue and prevent the binding of the peptide in an extended conformation [46].

In order to gain a better understanding on the role of this newly identified acidic patch on the PHD of BAZ2A/B towards the recognition of the H3 N-terminal tail, it was decided to perform some mutagenesis experiments on the proteins. Residues E1689 on BAZ2A and E1944 on BAZ2B were mutated to Gln and Lys in order to neutralize and/or invert the negative charge of the acidic wall. In order to compare the binding affinity of the H3 peptide toward the mutant proteins relative to wild type, I performed ITC experiments. Since the helical fold of the acidic wall was thought to be an essential structural feature to regulate the binding with H3 peptide, it was expected that a disruption of this wall could have led to a reduction of affinity between protein and peptide.

ITC were performed titrating peptide into proteins, in triplicate or duplicate. Peptide/protein concentrations were kept consistently identical through all the titrations in order to compare them appropriately.

Mutation to Gln of the acidic wall of BAZ2B led to ~8-fold reduction in affinity compared to the wild type, while the effect upon the same mutation on BAZ2A was less pronounced. Inversion of charge with mutation to Lys completely abrogated binding for the PHD of BAZ2B and reduced the affinity for the PHD of BAZ2A of ~2.5 fold (Table 1.3). Therefore in BAZ2B complete disruption of binding was reported while the acidic wall mutant in BAZ2A weakened binding. This could perhaps suggest that adjacent residues

could have compensating the effect on the mutation on the acidic wall. Indeed, it was further observed in the structure of the PHD BAZ2A in complex with H3 that the residue flanking E1689 is also an acidic residue D1688 and its side chain contributes to the formation of the pocket that accommodates the side chain of K4 onto the H3 peptide. In addition, despite the fact that the electron density for R8 on H3 peptide in our structure is not visible, superposition of the complex BAZ2A PHD - H3 10-mer peptide with the structure of MORF DPF in complex with H3K14bu [145] (PDB: 5U2J) suggests the formation of a salt bridge interaction between the side chain of R8 and the carboxylate group of D1688 (Figure 2.8).

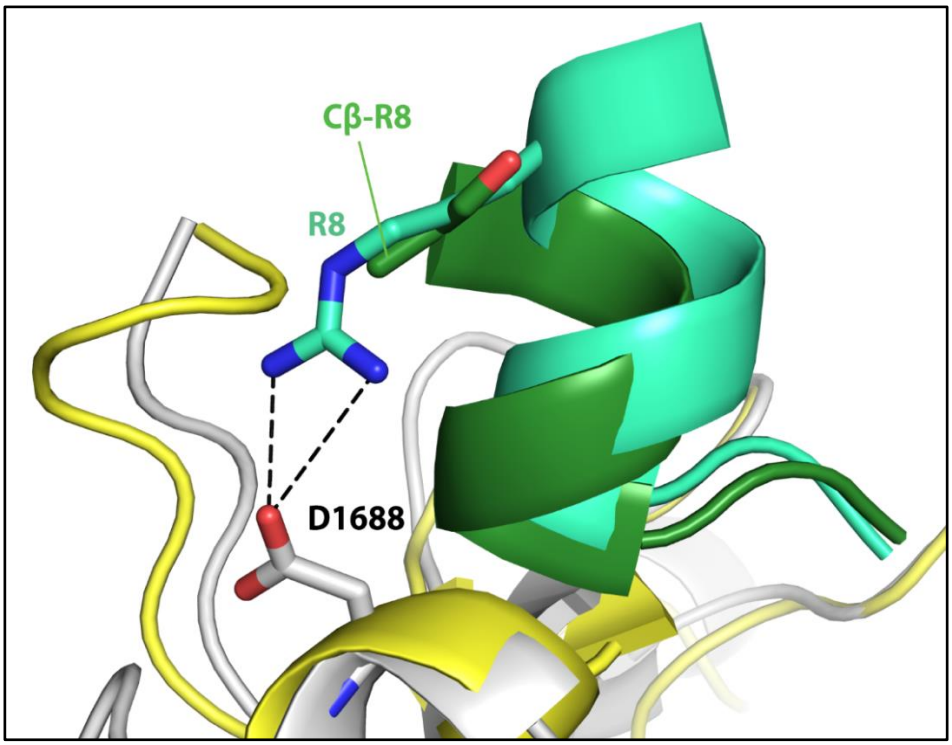


Figure 2.8. Superposition of BAZ2A PHD in complex with H3 peptide with the MORF DPF in complex with H3K14bu.

Superposition of the cartoon representation of PHD BAZ2A (grey): H3 10-mer (green) complex (PDB 5T8R) with the MORF DPF (yellow): H3K14bu (aqua green) complex. In stick the residues involved in the salt-bridge interaction.

Based on these observations, it was surmised that this residue could actually compensate the E1689Q mutation on BAZ2A. Consequently, I designed and expressed

another mutant carrying the double mutation D1688N/E1689Q, in order to neutralize the charge on the 3_{10} helix of BAZ2A. ITC experiments showed that the double mutation reduced the binding affinity of BAZ2A towards H3 10-mer up to ~17-fold.

Table 2.2 summarizes the thermodynamic binding parameters for the complex formation of H3 wt 10-mer with the mutants of BAZ2A/B PHD.

Figure 2.9 shows the ITC binding curves of the titration of H3 10-mer peptide into PHD zinc finger of BAZ2A/B. ITC titrations were performed and analysed as described in Chapter 6. Correct folding of the mutants was checked by 1D ^{1}H NMR spectra (A.7 in appendix).

PROTEIN	K_D (μM)	N	ΔH (KCAL/MOL)	TΔS (KCAL/MOL)	ΔG (KCAL/MOL)
BAZ2A PHD WT	48 ± 2	1.29 ± 0.01	-7.9 ±0.1	-2.0 ±0.1	-5.90 ±0.02
BAZ2A PHD E1689Q	83 ± 7	1.02 ±0.03	-9.6 ± 0.7	-4.0 ± 0.8	-5.57 ± 0.05
BAZ2A PHD E1689K *	116 ± 6	1.1 ±0.4	-7.6 ±2.5	-2.2 ±0.4	-5.37 ± 0.03
BAZ2A PHD D1688N/E1689Q	810 ± 50	1¥	-9.0 ± 1.2	-4.8 ± 1.2	-4.22 ± 0.04
BAZ2B PHD WT	47 ± 2	1.17 ±0.02	-8.3 ±0.2	-2.4 ±0.2	-5.90 ±0.03
BAZ2B PHD E1944Q	660 ± 60	1¥	-8.5 ±1.1	-4.1 ± 1.1	-4.34 ±0.06
BAZ2B PHD E1944K	>1000	N.D.	N.D.	N.D.	N.D.

Table 2.2. Summary of the ITC-based binding parameters for the formation of the complex of H3 WT 10-mer peptide with BAZ2A/B PHD mutants [46].

All the titrations were performed in triplicate, except where differently specified, and values are reported as the means ± s.e.m. Temperature of titration was 25°C.

N.D. = not determined

¥ N value was fixed at 1.

* titrations were performed in duplicate.

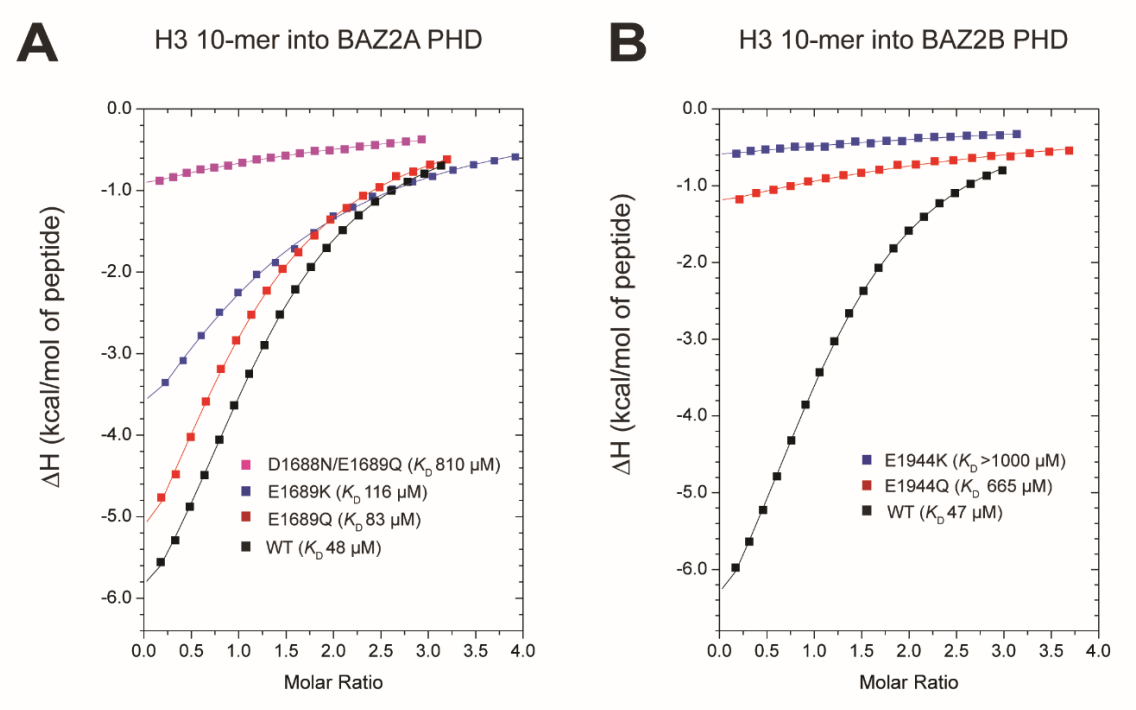


Figure 2.9. Acidic wall on BAZ2A/B PHDs and H3 N-terminal tail recognition.

ITC binding curves showing integrated ΔH (kcal/mol) values plotted versus the peptide/protein molar ratio. H3 10-mer peptide was titrated into wild-type and mutants of BAZ2A PHD (**A**) and BAZ2B PHD (**B**). Detailed legends for the mutations are reported in each panel.

2.5 STUDIES TOWARD ELUCIDATING THE BINDING MODE OF H3 TAIL WITH TANDEM PHD-BRD OF BAZ2A/B

Wang et al. in their work [146] showed that PTMs can affect the secondary structure of the histones, for example, acetylation was found to increase the α-helical content of the histone tail. In light of this and of the outcomes of the work herein presented, it might be theorised that also the secondary structure of the histone proteins could play a role in epigenetic regulation. Indeed, the structural rearrangement of the N-terminal histone tails, through either induction or destabilization of the helicity, could constitute a further layer of regulation for the epigenetic process, beyond or in cross-talk with PTMs recognition.

Indeed, there are examples in literature reporting how the helical folding of the H3 peptide is important for the simultaneous recognition of different region of the H3 histone by two adjacent reader domains. For example, this is the case of the PHD zinc finger and Tudor domain of UHRF1, where the helical conformation assumed upon the binding to the PHD is essential for the recognition of methylated K9 by the Tudor domain [147]. Another example is the simultaneous recognition of the unmodified K4 and acylated K14 on H3 by the double PHD zinc finger of MOZ or MORF; for both proteins the helical folding of H3 is found to be critical for the binding mode [58, 145] .

These observations, coupled to the results presented so far in this chapter, motivated further work aimed to shed new light and understanding of the combinatorial readout in the context of the combined PHD zinc finger and bromodomain tandem module (PHD-BrD) of BAZ2A/B proteins. Since it was previously shown that the adjacent bromodomain in both BAZ2A/B recognizes K14ac on H3 [45, 136], it is reasonable to postulate that the helical fold of H3 could play a role in the simultaneous recognition of both unmodified K4 and acetylated K14, respectively, within the same histone tail by PHD and bromodomain. This hypothesis would be consistent with data obtained so far showing that the two domains are not interacting [45] but would rather come close, facing each other, to bind the related PTMs without necessarily showing any interactions between themselves, as proposed by the SAXS data in antecedent work [45].

In order to investigate on the potential cooperative effect of the combinatorial readout of the tandem module of BAZ2A/B, I performed ITC experiments. Tandem modules of both proteins were expressed. BAZ2B PHD-BrD, upon purification, showed a good level of purity to be used in biophysical experiments, as reported in the SDS-page of figure 2. 10. In contrast, BAZ2A PHD-BrD showed to be more unstable and prone to degradation as evidenced by SDS-page gel and gel filtration profile, therefore this was not pursued for biophysical experiments.

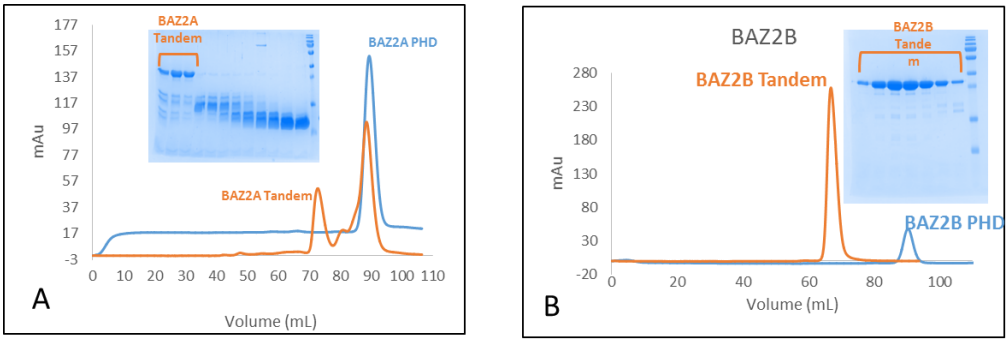


Figure 2.10. Gel filtration elution profile and SDS page gel of BAZ2A and BAZ2B PHD-BrD tandem.

A 20-mer peptide, resembling the H3 tail and carrying the acetylation mark on K14 (ARTKQTARKSTGG-Kac-APRKQL), was synthetized and purified (details in Chapter 6.3) and used as a tool for structural investigation. With these molecules in hand, I set to perform ITC experiments, titrating the peptide into the tandem protein (Chapter 6.4). A summary of the thermodynamic parameters measured by ITC is reported in Table 2.3 and raw data are shown in figure 2.11.

Protein	K_D (μM)	N	ΔH (kcal/mol)	TΔS (kcal/mol)	ΔG (kcal/mol)
BAZ2B Tandem	27.5 ± 1.3	1.62 ± 0.01	-6.5 ± 0.1	-0.25 ± 0.03	-6.2 ± 0.1

Table 2.3. Summary of thermodynamic binding parameters.

ITC-based binding parameters for complex formation between 20-mer H3K14ac peptide and BAZ2B PHD-BrD tandem. Error values reported on K_D, N, and ΔH are generated by the Origin program and

reflect the quality of the fit between the nonlinear least-squares curve and the experimental data.

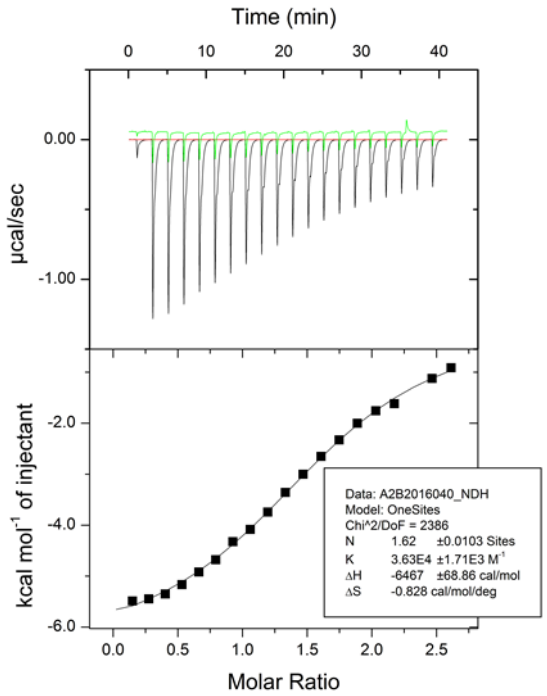


Figure 2.11. ITC binding curve if H3K14ac 20-mer vs BAZ2B PHD-BrD tandem.

ITC binding curve of H3K14ac 20-mer titrated into BAZ2B is shown in black and the relevant reference titrations (peptide into buffer) in green, in the upper panel. The integrated ΔH (kcal/mol) values are plotted versus the peptide/protein molar ratio and shown in the lower panel.

Stoichiometry value N ~ 1.6 can be explained considering the model in figure 2.12 and admitting a binding mode intermediate between the two states A and B, with a prevalence of the state B being populated at higher peptide concentration.

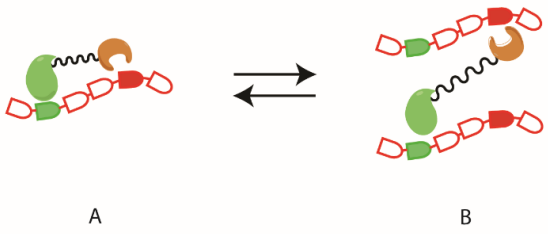


Figure 2.12. Two-state model of binding.

Cis (A) or *trans* (B) mode of binding of the PHD-BrD tandem.

In case of 1:1 binding event (case A) where the two reader domains from the same tandem protein simultaneously recognize hallmarks on the same peptide, an increase

of affinity due to a cooperative effect is expected. In our case the K_D value did not report any significant increase in binding potency compared to the binding between the single domain and the 10-mer peptide (K_D ~47 µM) [46] and more importantly no significant increase in the heat signal value was observed (ΔH ~-8.3 kcal/mol for the PHD and ΔH ~-6.5 kcal/mol for the tandem module).

The high energy required to bring close the individual domains PHD and BrD may account for the marginal energetic gain of the binding of the tandem to the H3K14ac 20-mer. Indeed, the long unstructured linker between the two may favour a binding in trans mode (case B in figure 2.12) and constitute an energetic barrier to the *cis* binding mode. Analysis of the profile of the ITC shows the presence of a shoulder at the end of the each peak that might be accounted for a structural rearrangement event happening during the binding of the PHD-BD tandem to the H3K14ac 20-mer peptide. The model built in figure 2.13 can be representative of the orientation that the two domain may assume upon binding.

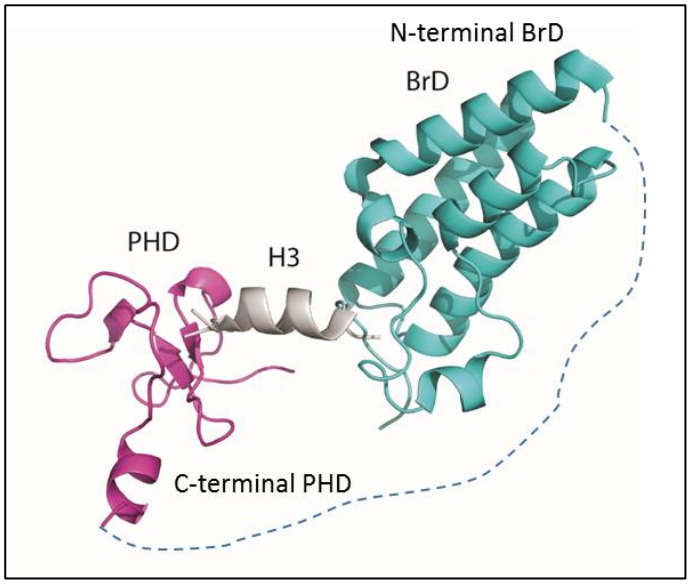


Figure 2.13. Proposed *cis* model for the binding of BAZ2B to the H3 histone tail.

Conversely, there are examples in literature of cooperative binding of PHD-BrD tandem (TRIM24 or TRIM33) showing enhanced affinity for the H3 peptide with the double marks compared to the binding of a single domain for the single-mark peptide [62]. All these observations pointed toward a prevailing of state B where the two reader domains

might separately bind the two marks on two different histone molecules. However, reasonable population of state A (*cis* recognition) at the early titration points of the curve cannot be excluded, indeed the *cis* recognition mode (figure 2.13) may still happen inside the cell.

In order to further gain insights on the binding mode, several crystallization attempts were performed with the goal to co-crystallize the complex PHD-BrD BAZ2B tandem - H3K14ac 20-mer, however these has remained inconclusive to date. The difficulty to co-crystallize this system could be justified assuming the promiscuous binding mode that leads to a heterogeneous complex intermediate between state A and B as in figure 2.11.

Nevertheless, the evidence to support the binding mode of the PHD-BD tandem of BAZ2A/B for histone H3 remains preliminary and requires further investigation.

2.6 DISCUSSION

This study centred on the molecular recognition mode of the H3 N-terminal tail by PHD zinc fingers of BAZ2A and BAZ2B. The major findings of my work was the solution of the crystal structure of the complex of BAZ2A PHD with H3 peptide which led to the discovery of the peptide helical binding mode. These findings motivated investigation of crystal structures available in the PDB of PHD domains with bound histone peptides. These analyses led to the realisation that PHD domains binding histone tails in a helical fold, as the case of the PHD of BAZ2A/B, are characterized by the presence of an acidic patch, containing Asp, Glu or both residues, on the loop or helical turn just before the first β-strand of the PHD [46]. The importance of this acidic patch was confirmed by mutagenesis experiments. I was able to show that indeed inversion or neutralization of the negative charge(s) led to a drop or complete loss of binding. Further work investigating artificial mutations on K4 or Q5, to Ala or Gly, to stabilize or disrupt, respectively, the peptide helicity, revealed an increase or decrease in binding affinity, consistent with the hypothesis that histone tail helicity plays a central role in the molecular recognition. These mutagenesis approaches on either the acidic wall patch of histone tail themselves could provide rapid and straightforward ways to identify other

reader domains that might recognize H3 in helical fold, prior to committing to detailed structural studies. The results of this study were recently published in *Biochem J.* [46]

The observations from this work have led to speculation that the helicity of the histone tail could represent an additional layer of regulation within the epigenetic processes. There are several studies which describe the importance of the helicity on H3 to guarantee the combinatorial readout of neighbouring domains. In this context, an investigation of the role of the secondary structure of the histone tail H3 upon recognition by the PHD-BD tandem of BAZ2B was conducted. The model here proposed seems to be compatible with a *trans* recognition mode of the histone tail on the nucleosome. According to this model, one protein, using two distinct reader domains (PHD and bromodomain) could bring closer two different nucleosomes through the binding of each domain to a distinct H3 tail. Although this data point toward a prevalent *trans* mode hypothesis of binding, the *cis* mode cannot be excluded, especially considering a recent work published by Kostrhon *et al.*, about the role of the linker on BAZ2B-histone interaction [148]. The authors found that the linker may negatively regulate the binding of the BrD to H3K14ac in the context of the tandem. This work has also opened another interrogative on the importance of the function of the long linker between two domains. Indeed, it is possible that the long linker region between the two reader domains acts as a scaffold to recruit partner proteins, leading to the formation of a larger complex. In this context, pull-down experiments followed by mass spectrometry analysis could help to elucidate this question.

Finally, the disclosure of this structural work has implications for drug discovery. Indeed, the PHD zinc finger still remain a class of reader domain poorly targeted and characterized by perceived lower druggability when compared to other reader domain families such as bromodomains. Taking advantage of the knowledge gained to date, the use of peptides able to stabilize the helical conformation e.g. stapled peptides [149, 150] or the use of a pharmacophore able to mimic the helical motif [151] in the histone pocket, could be a valid approach to design and develop chemical probes able to disrupt the Histone-PHD interaction in BAZ2A and BAZ2B PHD.

CHAPTER 3.
PROBING THE LIGANDABILITY OF THE PHD ZINC FINGER DOMAIN OF BAZ2A AND BAZ2B

3.1 MOTIVATION OF THE WORK

The aim of the work presented in this chapter is to explore the ligandability, i.e. their propensity to bind to small molecules, of two epigenetic reader domains: PHD of BAZ2A and BAZ2B. The ultimate goal of this project will be the development of chemical probes able to selectively bind these reader domains and which could be used as tools to investigate the activity and biological function of BAZ2A and BAZ2B proteins.

Targeting the PHD of BAZ2A/B has been shown to be challenging to date because of their shallow binding sites. Preliminary efforts in the lab with a fragment screening approach, involving high-throughput screening by biolayer interferometry (BLI) and differential scanning fluorimetry (DSF), had remained inconclusive. To view this in a different light, it was decided to adopt an alternative approach, performing a virtual screening and further biophysical validation, in order to aid assessment of the ligandability of these reader domains.

In this chapter, I will introduce the biophysical validation and structural characterization which I performed to validate some of the fragment hits previously identified by *in silico* screening. The two main techniques used were NMR and X-ray crystallography. The first one allowed a primary validation and estimation of a K_D of binding, while crystallography confirmed the hypothesis of binding formulated by NMR. Despite the high similarity between the two PHD proteins, some differences in the binding modes were observed.

3.2 CRYSTAL STRUCTURE OF THE TRIPEPTIDE 'ART' BOUND TO THE PHD OF BAZ2A

The alanine scan on the H3 10-mer peptide (discussed in Chapter 2) highlighted that the first three residues ART of the H3 histone tail are essential for the binding to the PHD domain of BAZ2A/B [46]. In order to elucidate the binding mode of this key N-terminal region of the H3 tail, I solved the crystal structure of the PHD of BAZ2A in complex with the tripeptide 'ART'. The complex was obtained by soaking overnight crystals of BAZ2A PHD in a solution containing the reservoir and supplemented with 20 mM ART peptide. The structure was solved following the same procedure used for PHD BAZ2A in complex with H3 10-mer peptide. Crystallographic data are summarized in Table 3.11.

Inspection of the structure led to the observation that three of the four protomers in the ASU (chains A, B and D) showed an extra electron density to fit the tripeptide ART. This is in contrast with the complex with H3 10-mer, in which only two chains (A and D) were binding to the peptide [46]. This can be rationalized as the short ART peptide occupies a smaller area and could have greater access than the H3 10-mer peptide through the solvent channels. Chain A and D are solvent exposed, while chain B and C are partially occupied by contacts with an adjacent asymmetric unit.

The analysis of the structure highlighted the hot spots of binding, reported in figure 3.1. The terminal amino group of A1 was found interacting through H-bonds and electrostatic interactions with the carbonyl of the backbone of P1714, E1715 and G1716. The carbonyl group of R2 was interacting with the N-H of L1693 while the N-H group of R2 is in contact with a structural water molecule, W1, located between the carbonyl of L1693 and the side chain of D1695. The hydroxyl group on the side chain of T3 is found to interact with the backbone carbonyl in E1689. The peptide is amidated at the C-terminus and the amidic nitrogen displays interaction with the backbone carbonyl of L1691 and D1688

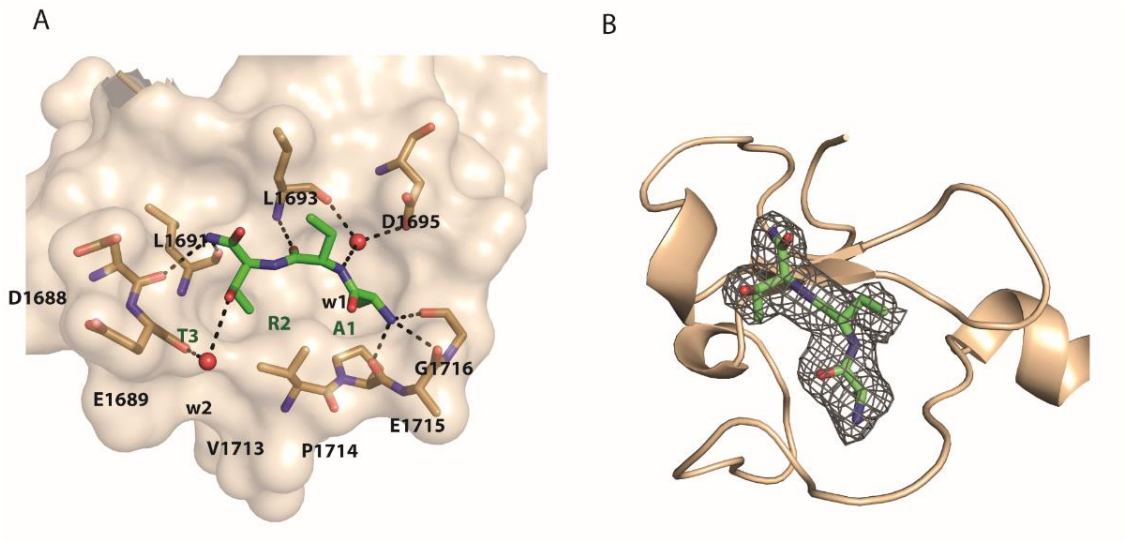


Figure 3.1 Complex of PHD of BAZ2A with H3 3-mer peptide.

(**A**) Surface representation of the crystal structure BAZ2A PHD (wheat) in complex with ART (green sticks) tripeptide. Highlighted in sticks protein residues interacting with the peptide. (**B**) Fo-Fc map for the 3-mer peptide is in grey and countered at 3σ.

The side chain of R2 seemed to be influenced by the crystal contacts present in the pocket. Because this residue was highlighted from the alanine scanning as an important residue for the interaction with the protein, the expectation was to find the guanidinium group interacting in a well-defined region of the binding site, making stable contacts. Instead, I observed that in the ART chains modelled in the histone pocket of protomers A and D, the electron density was clearly defined only up to the Cβ atom of the side chain, indicating that the rest of the residue does not make strong interactions. Moreover, in protomers A and D, it was observed that the carbonyl group of C1694 and C1697 make crystal contacts with the side chain of K1676 of chain C belonging to an adjacent ASU (figure 3.2 A). The presence of this contact in this position could generate electrostatic repulsion to the guanidinium group, which would be destabilized from the expected position interacting with the carbonyl from L1693 and the side chain of D1695. Protomer B has the histone pocket partially blocked, but despite this, clear electron density was observed and a molecule of ART tripeptide was built in. Modelling of the ART motif in this pocket showed that while the backbone was found in the expected position, the side chain of R2 was satisfactorily modelled in a way to make crystal contacts with the 3_{10} helix and β-strand of another protomer belonging to an adjacent ASU (figure 3.2 B). In particular, I observed H-bonds formation between the guanidinium

group of R2 and the carbonyl group of L1691 and D1688 of the histone pocket of an adjacent ASU. Because the presence of this crystal contact, no peptide was observed in chain C.

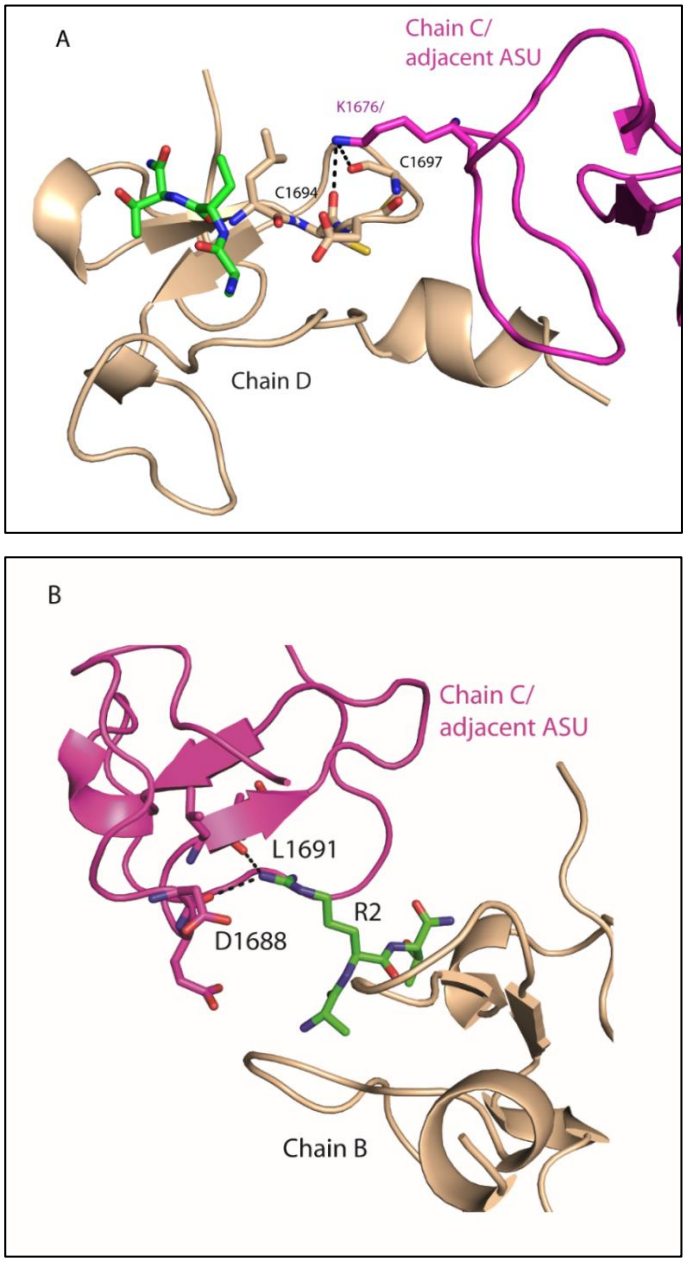


Figure 3.2. Crystal contacts between adjacent ASU.

(**A**) Chain D of one ASU is reported as wheat cartoon. In stick the two residues making crystal contacts with K1676 in chain C form the adjacent ASU. (**B**) R2 of the H3 3-mer peptide in complex with protomer B making crystal contacts with protomer C of the adjacent ASU.

3.2.1 Superposition of H3 10-mer complex with H3 3-mer complex

Superposition of the structures of BAZ2A PHD in complex with H3 10-mer and in complex with H3 3-mer (Figure 3.3) showed that the binding mode for the first three residues is retained in both structures, as expected, since they are critical to anchor the whole H3 tail to the PHD finger.

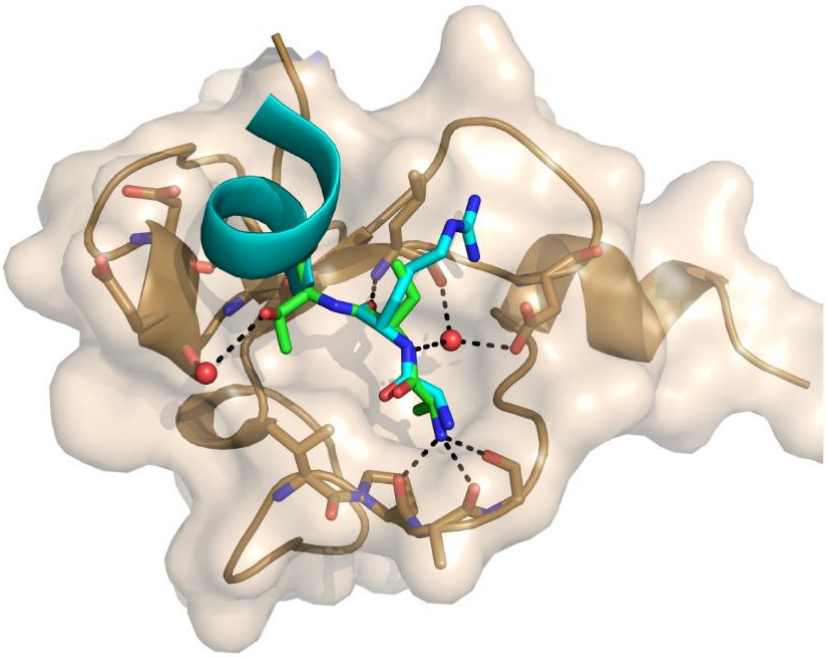


Figure 3.3. Superposition of BAZ2A PHD in complex with H3 10-mer peptide (PDB: 5T8R) and H3 3-mer peptide. Superposition of the crystal structure of BAZ2A PHD (wheat) in complex with ART (green sticks) tripeptide and H3 10-mer peptide (cyan). Highlighted in sticks protein residues interacting with the peptides.

The ART motif occupies an area that seemed adequate to accommodate fragment-like molecules. This observation would suggest that it might be possible to target the histone pocket of the PHDs with drug-like small molecules.

3.3 IDENTIFICATION OF DRUGGABLE POCKETS AND HITS FROM *IN SILICO* SCREENING

Using the FTMap software [2], Dr. Xavier Lucas identified two potential druggable pockets in the PHD of BAZ2A/B. The first pocket was the histone binding pocket [45, 46], perceived as druggable from the software, and here defined as pocket one. This is primarily acidic (figure 3.4). A second pocket was also found as druggable; this is located on the opposite side of the first pocket, on the rear of the protein (Figure 3.4). This second pocket is primarily basic and enriched with aromatic residues such as tryptophan and histidine.

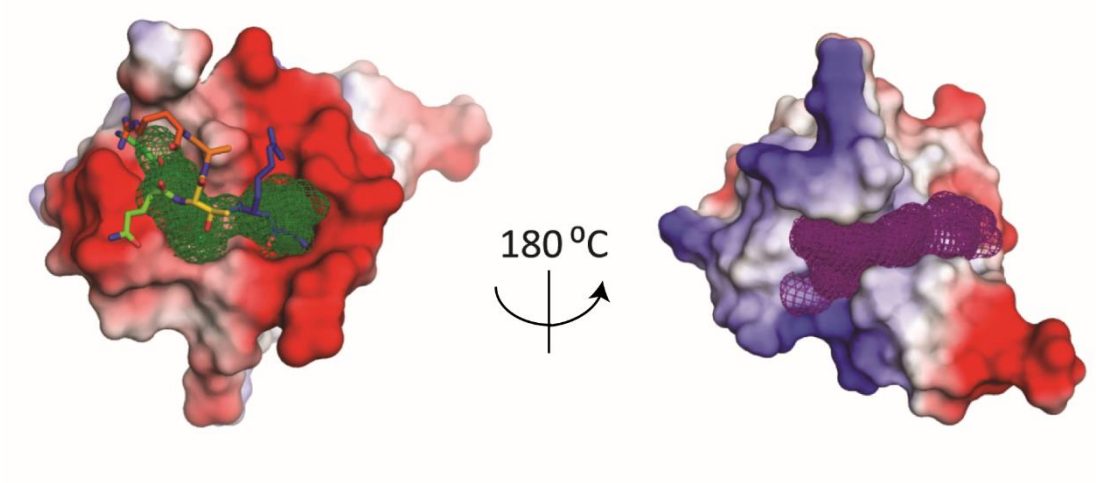


Figure 3.4. **Pockets perceived as druggable from FTMap on BAZ2A PHD.**

Interaction maps identified from FTMap software on PHD of BAZ2A. In green the FTMap of the histone pocket and in purple the FTMap of the second pocket.

In silico fragment screening was performed by Dr. Xavier Lucas using a virtual library of ~1,200 fragment-sized commercially available compounds. The final selection was trimmed by choosing only fragments that conformed to a set of criteria: calculated solubility in water (> 10 mM), chemical complexity (no more than two stereogenic centres), molecular weight (prioritizing lower number of heavy atoms) and fragments with opportunities for chemical modification.

Visual inspection of the virtual screening results led to 35 interesting candidates.

Amongst these, 15 fragments were selected from the in-house library, while additional four compounds from the virtual library were purchased (Table 3.1), for validation.

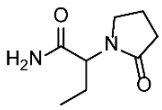
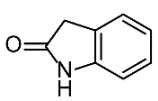
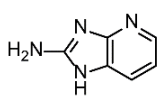
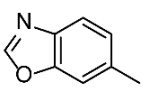
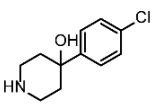
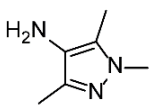
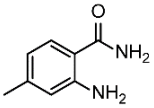
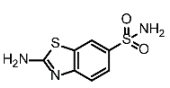
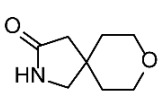
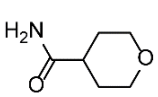
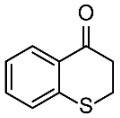
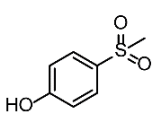
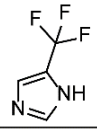
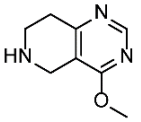
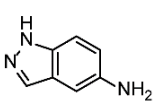
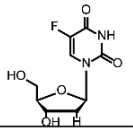
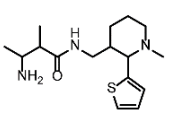
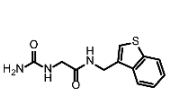
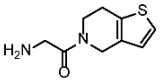
Fragment 1 	Fragment 2 	Fragment 3 	Fragment 4 	Fragment 5 	Fragment 6 
Fragment 7 	Fragment 8 	Fragment 9 	Fragment 10 	Fragment 11 	Fragment 12 
Fragment 13 	Fragment 14 	Fragment 15 	Fragment 16 	Fragment 17 	Fragment 18 
Fragment 19 					

Table 3.1. List of fragments selected by *in silico* screening.

Chemical structure of the 19 fragments selected from virtual screening.

3.4 NMR VALIDATION

3.4.1 (^{15}N-^{1}H)-HSQC validation

The pool of fragments identified in this first pilot screening was, in first instance, validated by (^{15}N-^{1}H)-heteronuclear single-quantum correlation (HSQC). This technique was chosen because it is one of the most sensitive to low-affinity interactions [152]. Moreover, it can provide information on the region of binding through chemical shift mapping. The small size of both proteins (6.5 kDa) allowed to generate high quality (^{15}N-^{1}H)-HSQC spectra of PHD of BAZ2A and BAZ2B, which resonances were previously assigned [46].

^{15}N -labelled protein was incubated with 5 mM of each fragment at final concentration of 1.25% (v/v) of d_6-DMSO. Spectra of proteins incubated with fragment were then overlaid with the spectrum of the *apo* form of the protein, obtained in presence of

matching 1.25% d_6-DMSO. Fragments showing chemical shift perturbation (CSP) in at least one peak, as in Figure 3.5, were considered binders. Those that did not show any relevant shift were discarded from the analysis. Fragment 4 was also discarded because not soluble in the presence of protein even at low concentrations as 500 µM.

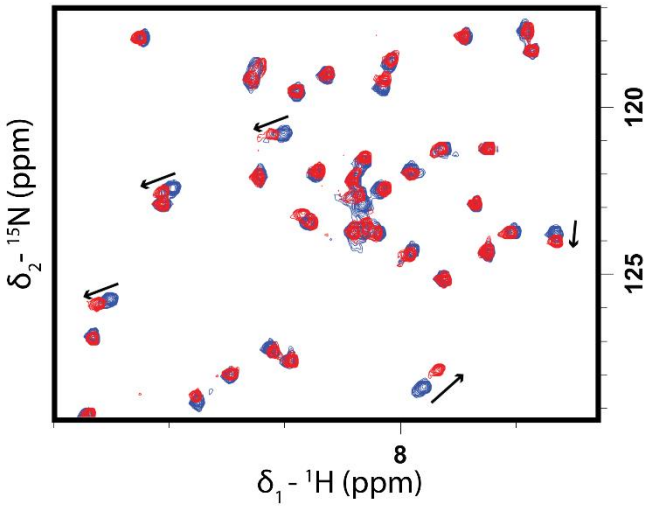


Figure 3.5. Example of CSPs induced upon fragment addition to PHD of BAZ2B.

Overlay of (^{15}N-^{1}H)-HSQC spectra recorded on the *apo* form of ^{15}N-BAZ2B PHD (blue) and upon addition of fragment (red). Arrows represent the direction of the shift of the peaks.

Amongst this small pool of fragments, seven compounds showed binding towards the PHD of BAZ2A and eight fragments showed binding to BAZ2B; most binders were common to the two PHDs, with just few fragments selectively binding BAZ2A or BAZ2B. Table 3.2 summarizes the results of this primary validation step.

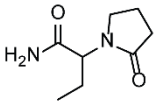
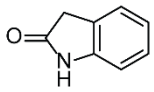
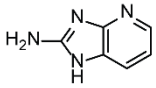
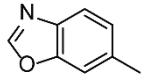
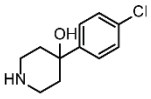
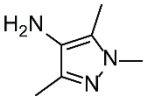
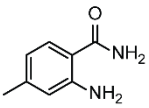
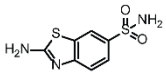
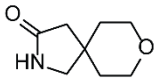
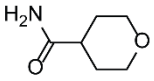
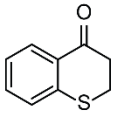
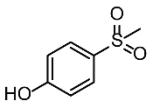
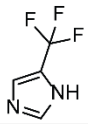
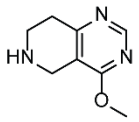
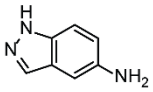
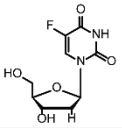
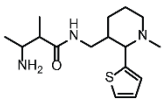
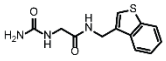
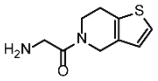
Fragment 1	Fragment 2	**Fragment 3**	Fragment 4	**Fragment 5**	Fragment 6
					
Fragment 7	**Fragment 8**	Fragment 9	Fragment 10	**Fragment 11**	Fragment 12
					
Fragment 13	**Fragment 14**	Fragment 15	Fragment 16	**Fragment 17**	**Fragment 18**
					
Fragment 19					
					

Table 3.2. List of fragments selected by *in silico* screening.

Chemical structure of the 19 fragments selected from virtual screening, highlighting those validated by HSQC. In red fragments found as communal hits for both PHDs BAZ2A/B. In blue are fragments binding only to PHD of BAZ2B, and in green fragment binding only PHD BAZ2A.

CSPs were plotted for each residue on a histogram and mapped on the structure of BAZ2A/B PHDs by colouring residues in red for the strong shifts and in orange for the medium shift. All those shifts reporting intensities lower than $\overline{\Delta\delta}$+ σ were considered as weak and coloured in grey. An example is shown in figure 3.6.

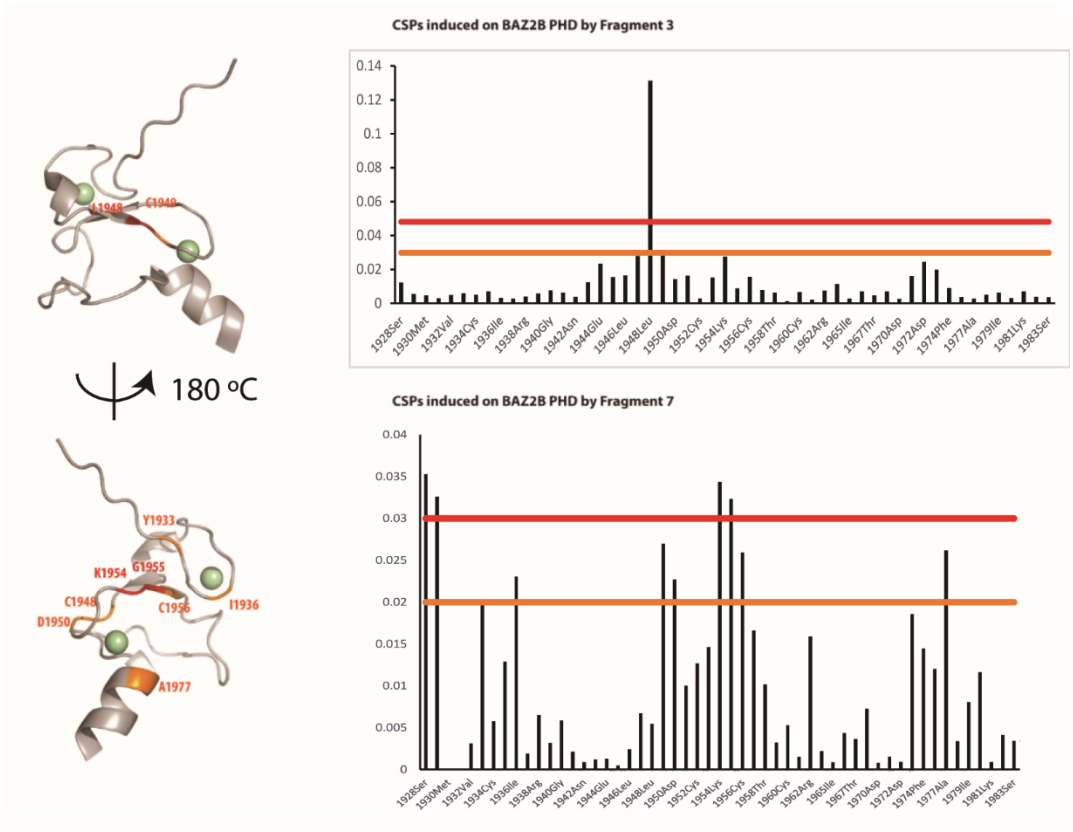


Figure 3.6. Examples of chemical shift perturbation induced on BAZ2B PHD upon addition of fragment.

Chemical shift differences induced on BAZ2B PHD upon addition of fragment 3 (upper panel) and 7 (lower panel) were plotted against the amino acid sequences of BAZ2B. The histograms, on the right, group the shifts according to their intensities: strong shift ($\Delta\delta > \overline{\Delta\delta} + 2\sigma$) and intermediate shift ($\Delta\delta > \overline{\Delta\delta} + \sigma$). The CSPs observed were mapped on BAZ2B PHD structures (PDB: 4QF3) by colouring residues with red and orange (respectively strong and medium shift). In the upper panel there are highlighted residues belonging to the histone pocket. In the lower panel, it is reported an example of shifts observed in the second pocket of the protein (180° rotation from the histone pocket).

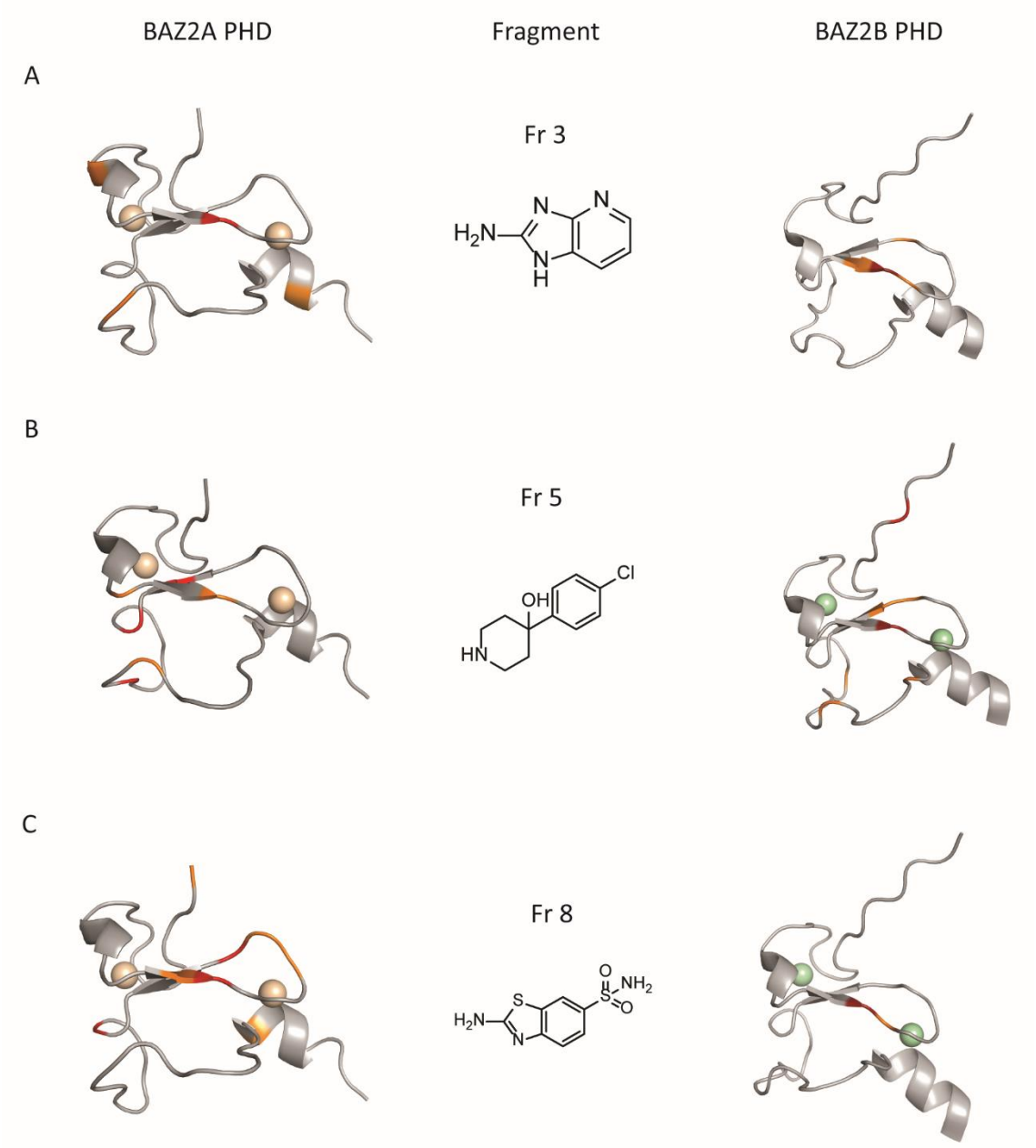


Figure 3.7 A-C. Heat maps of CSPs projected on BAZ2A/B PHD.

CSPs observed with the fragments reported in the middle column were mapped on BAZ2A (PDB: 4QF2, left) and BAZ2B (PDB: 4QF3, right) PHD structures by colouring residues with red and orange. Red is representative of strong shifts (Δδ > $\overline{\Delta\delta}$ + 2σ). Orange is representative of intermediate shifts (Δδ > $\overline{\Delta\delta}$ + σ). A, B and C respectively shows CSPs of BAZ2A and BAZ2B with Fr3, Fr5 and Fr8. The protein is illustrated with the histone pocket in frontal view.

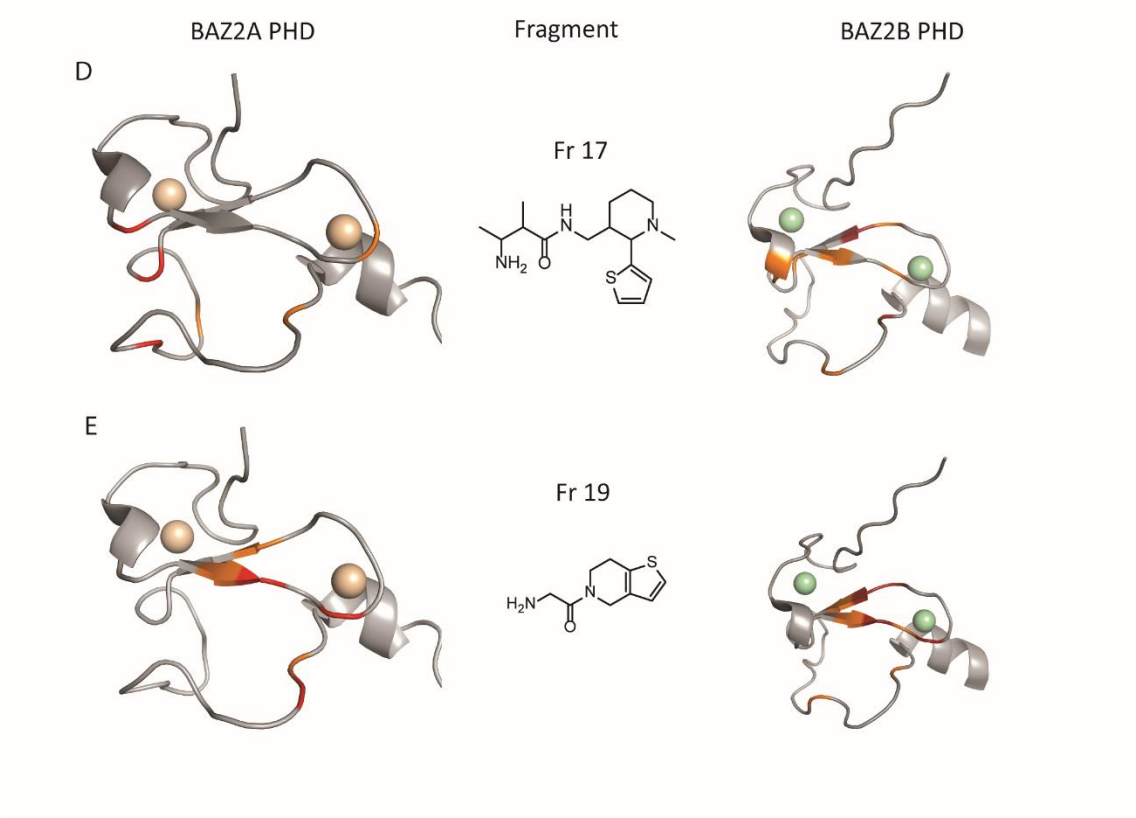


Figure 3.7 D-E. Heat maps of CSPs projected on BAZ2A/B PHD.

CSPs observed with the fragments reported in the picture (middle column) were mapped on BAZ2A (PDB: 4QF2, left) and BAZ2B (PDB: 4QF3, right) PHD structures by colouring residues with red and orange. Red is representative of strong shifts showing Δδ > $\overline{\Delta\delta}$ + 2σ. Orange is representative of intermediate shifts with Δδ > $\overline{\Delta\delta}$ + σ. Panel D and C respectively shows CSPs of BAZ2A and BAZ2B with Fr17 and Fr19. The protein is illustrated with the histone pocket in frontal view.

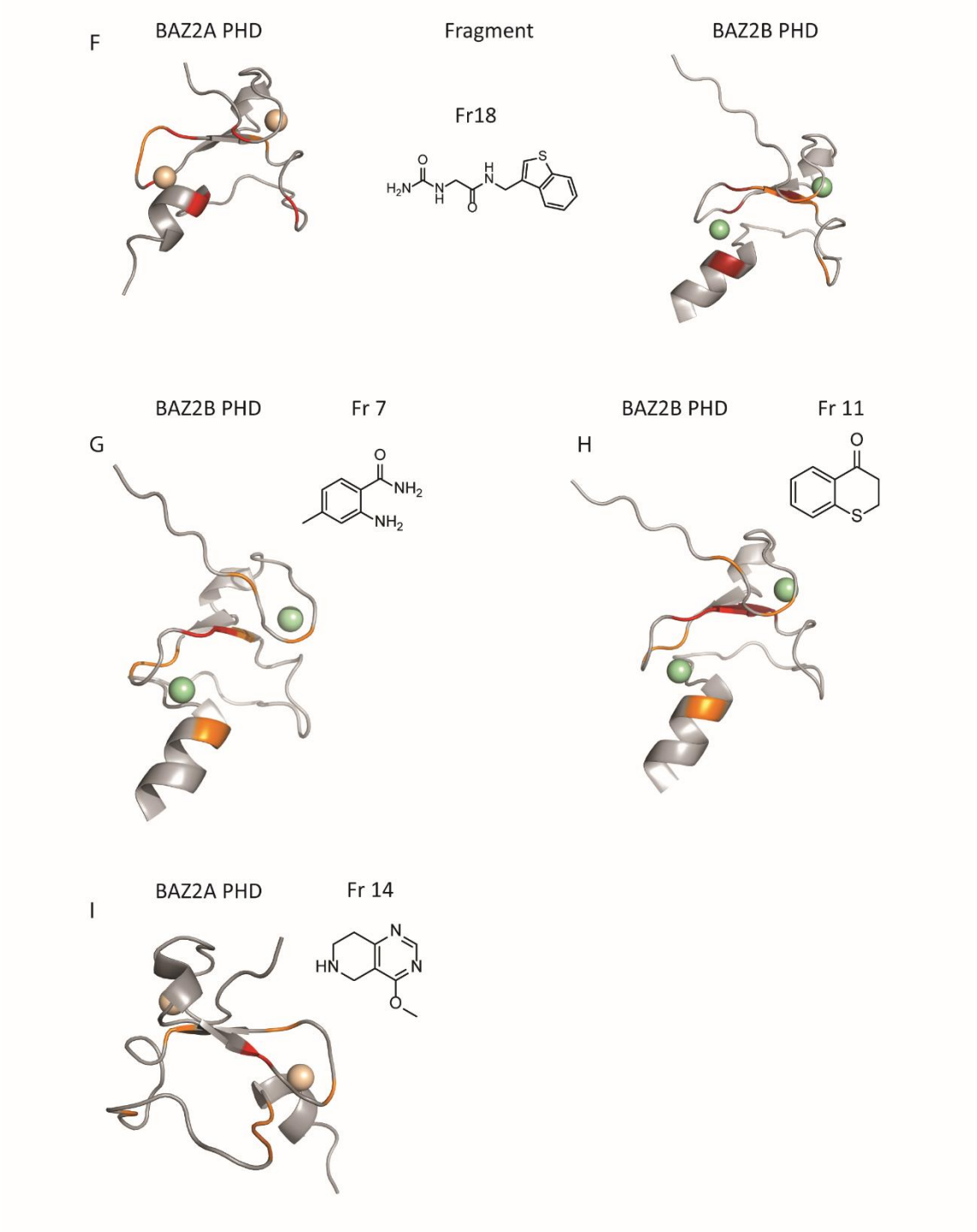


Figure 3.7 F-I. Heat maps of CSPs projected on BAZ2A/B PHD.

CSPs observed with the fragments reported in the picture (middle column) were mapped on BAZ2A (PDB: 4QF2, left) and BAZ2B (PDB: 4QF3, right) PHD structures by colouring residues with red and orange. Red is representative of strong shifts showing Δδ > $\overline{\Delta\delta}$ + 2σ. Orange is representative of intermediate shifts with Δδ > $\overline{\Delta\delta}$ + σ. In panel F, G and H the rear view of the protein is reported. In F, CSPs of BAZ2A and BAZ2B interacting with fragment 18. In G and H, it is reported only BAZ2b PHD interacting respectively with Fr **7** and Fr **11**. In panel I, frontal view of BAZ2A PHD with CSPs from the interaction with Fr **14**.

Fragment 3 (1H-imidazo[4,5-b]pyridine-2-amine) and 8 (2-amino-1,3-benzothiazole-6-sulfonamide).

A first difference between the two proteins raised up from the different intensities of the CSPs of Fr **3** and Fr **8**. CSPs of both fragments were much more pronounced in BAZ2B than BAZ2A.

Interestingly, Fr **3** and Fr **8** present a highly similar scaffold. They both contains two condensed aromatic systems, respectively a 1H-imidazo[4,5-b]pyridine and a benzothiazole, carrying an amino group on the azole derivative. Remarkably, a previous study, reported by Miller et al. [57], identified the same 2-amino benzothiazole fragment as a binder of Pygo PHD. This amino group, in this particular position, seems to be the driving force of the binding. Indeed, it was noticed that also Fr **15** contains a similar scaffold constituted of two condensed aromatic ring (an indazole) which also carry an aromatic $–NH_2$ group but on the phenyl ring. Fr **15**, conversely to Fr **3** and Fr **8**, did not show any CSPs.

Fr **8** contains a sulphonamide motif which is capable of binding to metals. This functional group of the molecule could potentially interfere with the PHD zinc finger chelating the zinc ions within the protein. This possibility was eventually discarded since the CSPs, projected on both PHDs, display a defined cluster of residues in the region of the histone pocket. Nevertheless, in BAZ2A PHD, some of the residues adjacent to one of the zinc ions are characterized by intense shifts, suggesting potential interference in that region of the protein (figure 3.7 C). X-ray crystallography will prove invaluable for elucidating the binding mode of this fragment.

Fragment 5 (4-(4-chlorophenyl)-4-hydroxypiperidine)

Fr **5** was one of the ligand that produced the most intense shifts in both proteins, although the mapping appears to be different between the two PHDs (Figure 3.7 B). It contains amine and hydroxyl groups that could act as HBDs and a clorophenyl ring which could be involved in aromatic stacking. From *in silico* modelling, Fr **5** is supposed to bind to the histone pocket and the CSPs reported for BAZ2B PHD seems to be in agreement with the prediction. For BAZ2A, CSPs do not identify a specific cluster of binding and some peaks were also disappearing upon titrations. This observation could indicate a level of non-specific binding.

Fragment 7 (2-amino-4-methylbenzamide) and fragment 11 (thiochroman-4-one).

Fr **7** and Fr **11** selectively showed CSPs only for the PHD of BAZ2B. Mapping of the shifts identified as potential region of binding the second pocket on the rear of the protein (Figure 3.7 G and H). The similar scaffold between the two compounds may account for the analogous response detected by HSQC. Indeed, both fragments contain a phenyl ring (likely involved in van der Waals interactions and π-π interactions) and a carbonyl group that could act as hydrogen bond acceptor.

Fragment 14 : 4-methoxy-5,6,7,8-tetrahydropyrido[4,3-*b*]pyrimidine

Fr **14** showed weak shifts only for PHD of BAZ2A, located in the region of the histone pocket (Figure 3.7 I), while the *in silico* prediction reported potential binding also for BAZ2B PHD.

Fragment 17: 3-amino-2 methyl-N-((1-methyl-2-(thiophen-2-yl)piperidin-3-yl)methyl)butanamide

This molecule presents a larger molecular weight compared to the other fragments tested. It presents a peptidomimetic scaffold and for this reason was predicted to bind to the histone pocket. In BAZ2B showed a cluster of shift related to the histone pocket, although a very intense shift is also reported for K1955 on the rear of the protein (Figure 3.7 D, right side). In BAZ2A, the sequence of shifts registered does not define a proper pocket, making any hypothesis of binding difficult (Figure 3.7 D, left panel). However, it must be noted that this molecule presents four chiral centres that will generate a racemic mixture of 16 diastereoisomers. In this case, we might have different scenarios where more than one species is potentially binding the protein (generating non-specific binding that could account for the CSPs trend observed in BAZ2A) or the binding can be specifically determined by one single diastereoisomer.

Fragment 18: N-(benzo[b[thiophen-3-ylmethyl)-2-ureidoacetamide

Fr **18** also presents a peptidomimetic scaffold. It contains many possible groups for binding, including three amines, two carbonyls, and a benzothiophene group. In both proteins the CSPs target the second pocket located on the back of the protein (figure 3.7 F). In this case, we could assume that the binding might be driven from the

benzothiophene group.

Fragment 19: 2-amino-1-(6,7-dihydrothieno[3,2-c]pyridine-5(4H)-yl)ethan-1-one

This fragment has a small region with peptidomimetic characteristic plus a thiophene group that could be useful to catch hydrophobic interactions. By HSQC it was found to bind both proteins in the histone pocket, as expected. Residues reporting shifts are the hotspots of binding for H3 peptide, on the first β-strand of the PHD [46], as shown in Figure 3.7 E.

Impurities associated with fragments can often cause binding interference, in particular there are works from the literature suggesting how metals coming from the synthesis may influence the hits of binding [134 153]. The effect of metals can be prevented using in the assay chelating agents as EDTA. Unfortunately, this precaution was not achievable in our NMR validation experiments since the proteins under investigation contain two structural zinc ions and the addition of EDTA would have caused denaturation of the domains.

3.4.2 Validation by ligand-observed NMR

Considering the low MW of the target protein, it was deemed interesting to investigate if it was possible to monitor protein-ligand interactions by ligand-observed NMR techniques. For this study, I tested some of the fragments already validated by HSQC: Fr **3**, Fr **5**, Fr **7**, Fr **8** and Fr **11**. As model of study I chose BAZ2B PHD because this protein gave more pronounced CSPs with fragments in the HSQC experiments.

WaterLOGSY showed low positive signal upon protein addition for Fr **3**, Fr **5** and Fr **7** (figure 3.8), while for Fr **8** and Fr **11** no change in response was detected. In contrast, in the STD spectra recorded, for all the fragments, upon protein addition, no fragment signal was detected. CPMG spectra were initially recorded with relaxation delays of 60 ms. In this condition no relevant signals were detected. A trial of different relaxation time was attempted. Six different spin-lock time were applied: 40, 120, 160, 320, 640 and 1280 ms. Majority of the fragments did not show any response despite the changes. Instead Fr **3** and Fr **8** generated sensible CPMG spectra only with specific spin-lock time.

Fr **3** with delay at 320 and 640 ms while Fr **8** only at 160 ms (figure 3.8).

Despite these changes, responses obtained were not convincing to state the viability of these techniques towards detecting ligand binding in our model study. So, no further investigation by ligand-observed NMR was conducted.

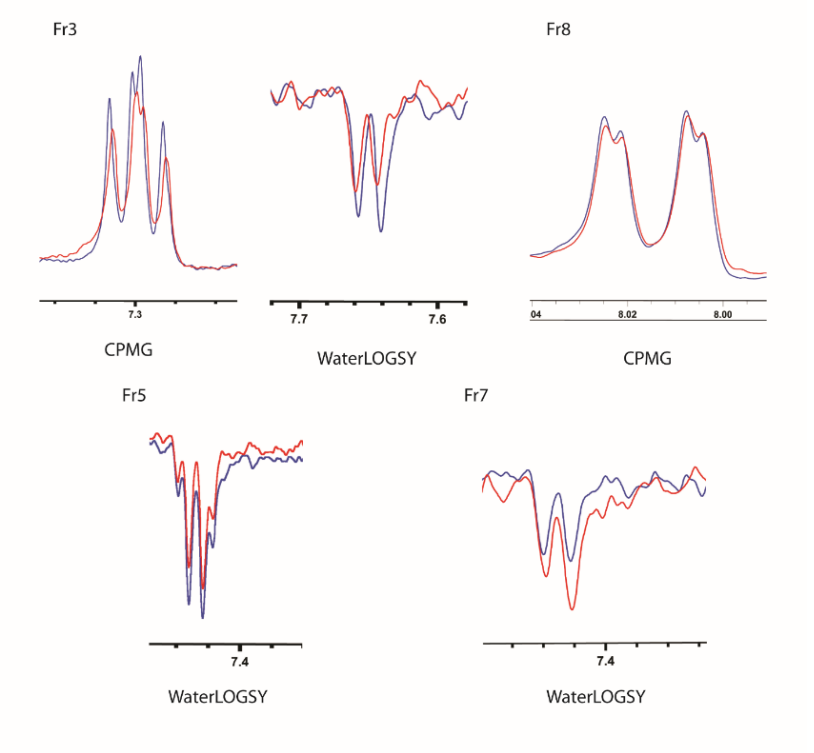


Figure 3.8. Ligand-observed NMR performed for PHD BAZ2B with fragments.

CPMG and WaterLOGSY spectra of fragments that gave positive responses. In blue fragment alone, in red fragment upon protein addition.

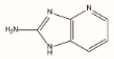
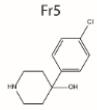
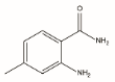
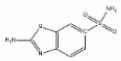
Fragment	STD	CPMG	WaterLOGSY
Fr3 	✗	✓	✓
Fr5 	✗	✗	✓
Fr7 	✗	✗	✓
Fr8 	✗	✓	✗
Fr11 	✗	✗	✗

Table 3.3. Ligand-observed NMR test.

Five selected fragments, validated by protein-observed NMR, tested by ligand-based NMR. Crosses indicate negative response detected and ticks indicate positive response detected.

3.5 DIFFERENTIAL SCANNING FLUORIMETRY (DSF)

DSF was next used as an orthogonal technique to confirm binding already observed by HSQC-NMR. Prior to testing the fragment by DSF, the best conditions to perform the assay were evaluated.

3.5.1 Identification of screening conditions

Previous experiments in the lab had tested the stability of the PHD BAZ2A/B in different buffers at different pH. It was chosen to perform the fragment validation using a buffer condition where the T_m was as close as possible to 37 °C, in a way to resemble conditions close to the physiological one. Buffer containing 100 mM MES at pH 6.0 was considered suitable. The buffer concentration was kept high in order to ensure good buffering capacity in the presence of high fragment concentrations. In addition, the effect of DMSO and NaCl on protein stability was evaluated. A first screening was performed with

both proteins PHD BAZ2A/B testing two different salt concentrations and different percentage of DMSO up to 5% v/v (figure 3.9).

It was observed that the increase of DMSO did not affect the stability of the proteins, indeed, the Tm remained in the same range. The increasing concentration of salt showed a more pronounced variation of the Tm, with protein stabilisation already observed at 50 mM NaCl (figure 3.9). Eventually, it was chosen to perform the screening at 100 mM MES pH 6.0, 5% (v/v) DMSO and 50 mM NaCl, in order to guarantee stability of the protein and high solubility of the fragments at all the concentrations, without increasing the ionic strength which could prevent the binding of the fragment to the pocket (figure 3.9).

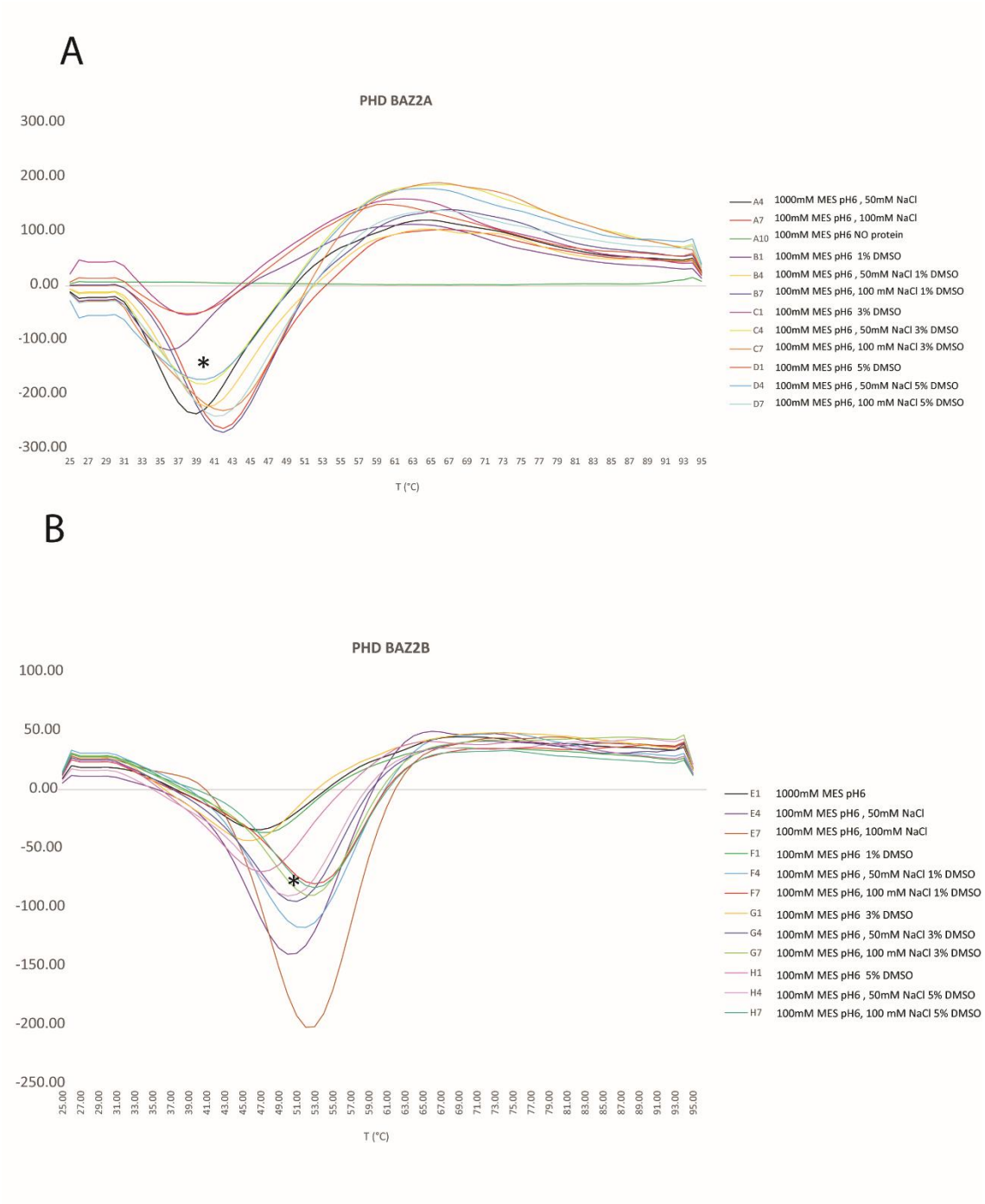


Figure 3.9. Optimization of DSF conditions prior to performing the fragment screening.

Conditions screened for the optimization are listed on the right side. Derivative of the melting profile of PHD of BAZ2A and PHD of BAZ2B are reported. The asterisk individuated the conditions selected for the screening: 100 mM MES, 50 mM NaCl and 5% DMSO.

3.5.2 Fragment screening

Eight fragments identified by HSQC NMR as positive hits were tested by DSF. The H3 10-mer peptide that is known to bind the PHDs was used as positive control and tested at 500 µM. Each fragment was tested at three concentrations: 3, 5 and 15 mM.
Table 3.4 reports the results of the screening.

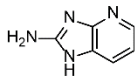
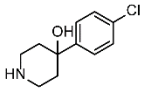
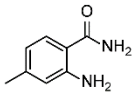
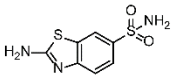
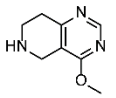
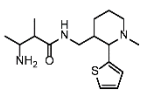
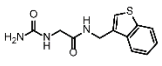
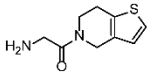
Fragment		ΔTm (°C) PHD BAZ2A	ΔTm (°C) PHD BAZ2B	[Compd]
H3 10 mer wt		14 ± 0.3	12 ± 0.6	500 µM
Fr3		ND	ND	3 mM 5 mM 15 mM
Fr5		1.3 ± 0.5 3.0 ± 0.7 9.8 ± 0.5	1.4 ± 0.6 2.3 ± 0.9 5.9 ± 1.4	3 mM 5 mM 15 mM
Fr7		NT	-0.9 ± 0.7 -1.9 ± 0.9 ND	3 mM 5 mM 15 mM
Fr8		ND	ND	3 mM 5 mM 15 mM
Fr14		1.2 ± 0.7 1.5 ± 0.3 1.9 ± 0.5	1.9 ± 0.9 2.2 ± 0.6 ND	3 mM 5 mM 15 mM
Fr17		2.2 ± 0.3 NT NT	4.9 ± 0.9 NT 13.9 ± 1.5	3 mM 5 mM 15 mM
Fr18		-1.7± 0.6 ND ND	ND	3 mM 5 mM 15 mM
Fr19		-1.4 ± 0.3 -0.9 ± 0.4 NT	0.5 ± 0.8 -2.2 ± 0.6 0.6 ± 0.6	3 mM 5 mM 15 mM

Table 3.4. Summary of ΔTm calculated by DSF.

Summary of the ΔTm calculated by DSF for each protein upon addition of 3, 5 and 15 mM of fragment. On the left the chemical structure of the fragments tested and already validated by HSQC.

The majority of the fragments showed a stabilization of the protein with positive ΔTm, increasing at increasing fragment concentration, as expected for genuine reversible 1:1 binding.

Fr **14**, which was expected to bind only BAZ2A PHD based on the NMR data, was in fact found to stabilize both proteins by DSF.

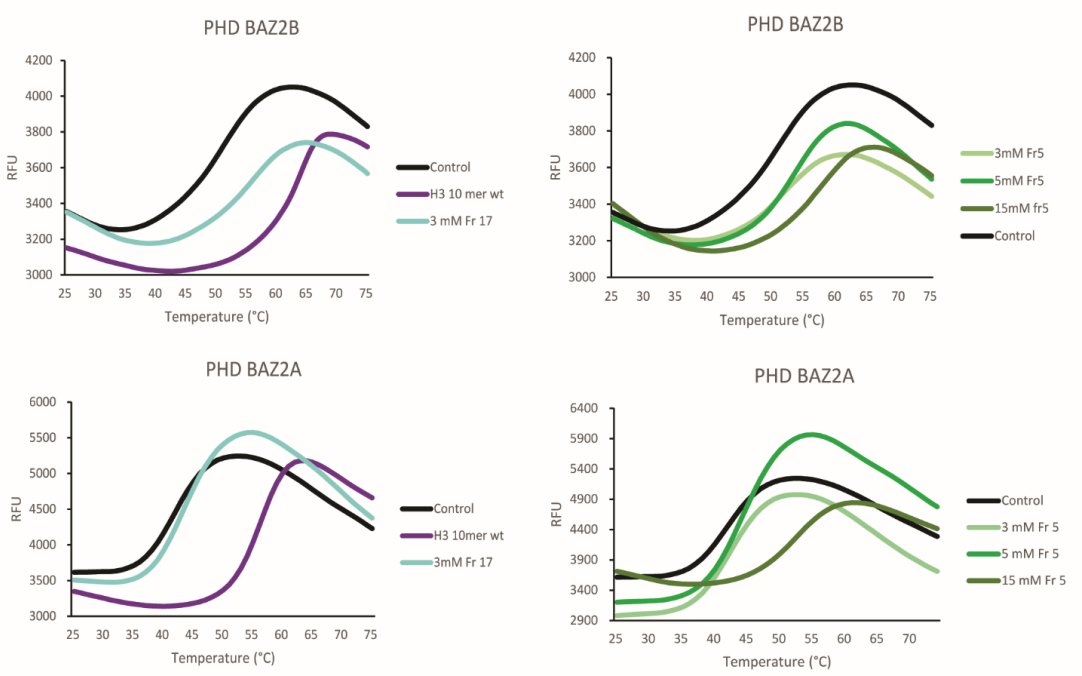


Figure 3.10. DSF melting curves.

DSF melt curves of BAZ2A and BAZ2B PHDs (black) in presence of H3 10-mer peptide (purple) and 3 mM Fr **17** (cyan) on the left side. On the right side, DSF melting curves of BAZ2A/B PHD in presence of increasing concentrations of Fr **5**. Data are reported as relative fluorescence units (RFU) *vs* temperature (°C).

Examples of some DSF melting curves are reported in figure 3.10. In black the protein alone, in purple the protein in presence of the H3 10-mer peptide (used as control) and in cyan the protein in presence of the Fr **17**, in a shifted position intermediate between the protein alone and the complex protein- peptide, as expected. On the left are shown thermal melting profiles of the protein alone and in presence of the three increasing concentration of a fragment that stabilize the protein. In this case, the observed shift increases with the increasing of the fragment concentration. The highest shift is always

registered in presence of 15 mM of fragment (darker green).

Fr **7** was tested only for the PHD of BAZ2B since it did not show any shifts for BAZ2A PHD in the previous NMR validation. As reported in Fig. 3.11 and in Table 3.4, it destabilizes the protein with a ΔTm of -0.9 at 3 mM fragment and -1.9 with 5 mM fragment. It has been proven, from previous studies reported in literature, that also ligands that destabilize the protein can be real binders [133]. Fr **18** also showed a negative ΔTm for PHD BAZ2A while for BAZ2B no detection was possible. Fr **18** precipitated at concentration higher than 3 mM, generating high fluorescence already at low temperatures, and not detectable signal.

Figure 3.11 reports melting curves generated by destabilization of the protein upon binding with Fr **7**. DSF curves appear to be flatter than the curves of the protein alone.

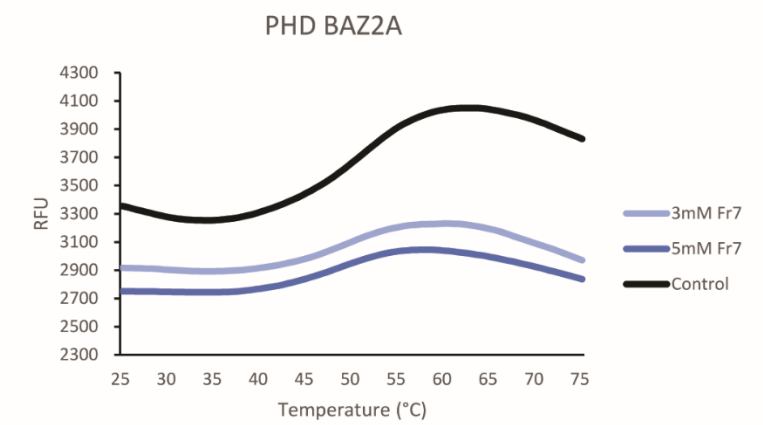


Figure 3.11. DSF melting curves with Fr 7.

DSF melt curves relative to the destabilizing Fr **7** in complex with BAZ2B PHD. Relative fluorescence (RFU) reported against temperature. In black curve of the protein alone, in light blue protein in presence of 3mM Fr **7** and in dark blue protein in presence of 5 mM Fr **7**.

Fr **3** and Fr **8** did show intrinsic fluorescence signal in any of the three concentrations screened. This effect could be explained by either aggregation of the protein in presence of fragment, aggregation of fragment itself or possible interactions between the fragment and the SyproOrange probe.

3.6 DETERMINATION OF BINDING CONSTANT K_D

3.6.1 Estimation of binding constant by (^{15}N-^{1}H)-HSQC

One of the crucial step during fragment screening in drug discovery is the determination of binding constants. The purpose is to rank the fragments for further optimization and lead selection.

In this study, it was decided to use Chemical Shift Perturbation (CSP) analysis to determine the binding constants since the domains studied have a low molecular weight and yielded to good quality HSQC spectra. Given that fragments are generally weak binders (K_D >10^{-5} M), our expectation is to observe shifts in the fast exchange regime (k_{ex} >> Δω). This means that we should be able to detect the binding within the NMR scale time and following the variation of the chemical shifts during addition of increasing concentrations of compound, it would be possible to extrapolate information about the binding constant.

Conditions of the assay were changed compared to the NMR validation step. Buffer concentration was increased up to 50 mM $H_2PO_4^-$/HPO_4^- in order to improve the buffering capacity also at high fragment concentration (till 5 mM). No salt was added to limit potential screening away of electrostatic interactions. Some of the fragments showed higher solubility and were directly dissolved in the NMR buffer; other fragments were previously dissolved in DMSO. It was noticed that up to 5% DMSO no substantial differences in the chemical shift were observed in the HSQC spectrum compared to the apo form. Consequently, as reference was used the spectrum of the apo form of the protein.

Due to the low ^{15}N -labelled proteins availability, titrations were performed adding known ligand volumes to the protein samples, in order to limit the protein consumption. The added volume was kept low (no more than 5 µL), using a high fragment stock concentration, to minimize the errors on the fitting due to the protein dilution. Protein concentration was kept at 100 µM which is the lowest concentration still yielding good quality spectra, with sharp and defined peaks. Fragments were titrated in a range between 0.5 and 5 mM, according to the solubility limits of the compounds. Concentrations higher than 5 mM were not used to prevent potential aggregation either

of the protein or of the compound, to avoid changes in pH of the solution that may alter the signals and confuse the analysis as well as other artefacts associated with high compound concentrations.

All the fragments identified by HSQC were titrated versus the protein where CSPs were previously observed at the singleton screening.

The titration pattern observed, confirmed that most fragments follow the fast exchange regime, as expected, since the affinities are in the mM range. Indeed, the majority of the peaks, affected by binding, move smoothly from the unbound to the bound form, with the exception of few peaks disappearing or broadening (potential indication of non-specific binding or intermediate exchange).

The analysis of the spectra and K_D estimation were performed using the CcpNMR software [154], which allows to calculate the weighted shift changes Δδ at different ligand concentration and K_D value for each residue of the protein. Indeed, for each fragment, it is possible to extrapolate a K_D for each single residue of the protein, but only those residues giving meaningful shifts ($\Delta\delta > \overline{\Delta\delta} + \sigma$) were accounted for the calculation of the overall K_D, since involved in binding. The final K_D is then calculated, for each fragment, as mean of K_D values derived from CSPs of those residues showing $\Delta\delta > \overline{\Delta\delta} + \sigma$. An example representative is reported in Figure 3.12. In table 3.5 there is a summary of the K_D of binding, estimated by HSQC, for the hits previously validated.

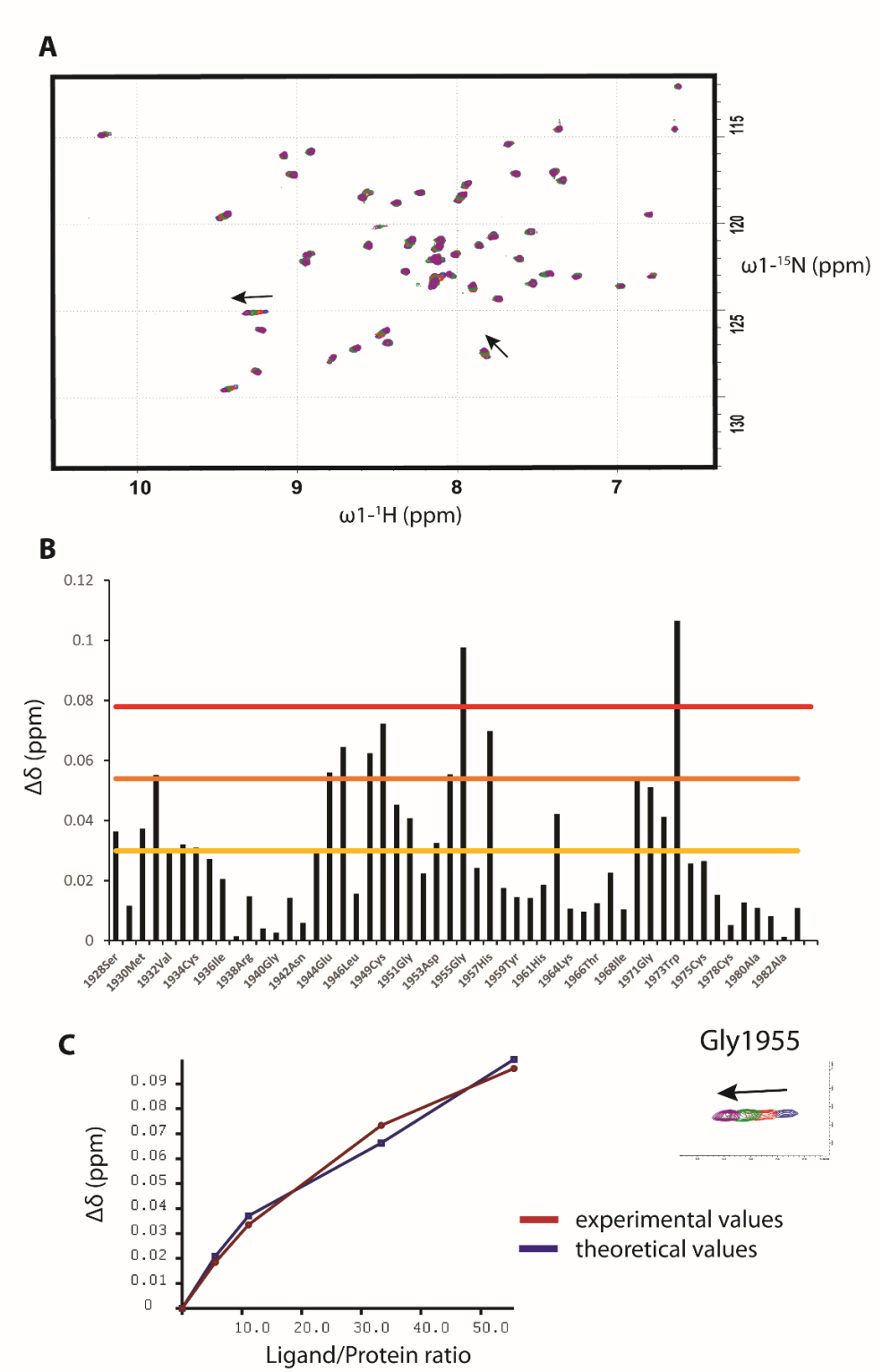


Figure 3.12. Fr 17 binding to PHD of BAZ2B.

(A) Overlay of the spectra of PHD of BAZ2B upon addition of different fragment concentrations. (B) Histogram of CSPs plotted for each residue of the protein. (C) Plot of Δδ (ppm) versus ligand/protein ratio of the residue Gly1955 of BAZ2B. In red the experimental values, in blue the theoretical values. Chemical shifts reported in different colour at different concentrations: blue, no fragment; red, 1 mM Fr **17**; green, 3 mM Fr **17** and purple, 5 mM Fr **17**.

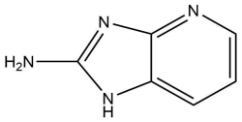
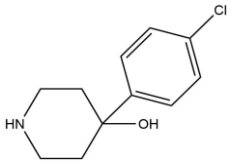
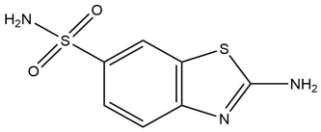
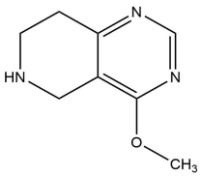
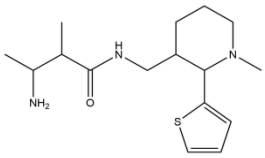
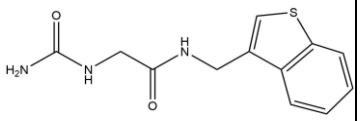
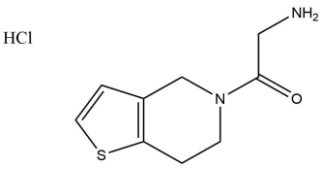
Fragment structure	K_D (mM) BAZ2A PHD	K_D (mM) BAZ2B PHD
Fr 3 	NT	4.3 ± 1.3 (LE 0.36)
Fr 5 	4.7 ± 1.9 (LE 0.23)	>16 (LE 0.17)
Fr 8 	NT	4.4 ± 0.7 (LE 0.23)
Fr 14 	9.3 ± 1.6 (LE 0.23)	NT
Fr 17 	10.8 ± 2.1 (LE 0.13)	3.4 ± 0.3 (LE 0.16)
Fr18 	9.2 ± 2.6 (LE 0.15)	2.3 ± 0.5 (LE 0.2)
Fr 19 	NT	9.4 ± 1.8 (LE 0.2)

Table 3.5. Summary of K_D estimation by HSQC NMR and ligand efficiency (LE).

K_D are calculated as mean ± s.e.m. of single K_D extrapolated from CSPs. In brackets, ligand efficiency in (kcal/mol)/heavy atom. NT = not tested.

In the first HSQC validation screening, Fr **3** and Fr **8** were individuated as possible binders of both proteins. During the titration it was observed that, while for BAZ2B increasing Δδ were observed at increasing of fragments concentration, for BAZ2A only one residue was showing Δδ, displaying also not relevant changes upon titration. Consequently, Fr **3** and Fr **8** were not further investigated for K_D evaluation towards BAZ2A.

Fr **7** and Fr **11** showed toward BAZ2B a trend similar to Fr **3** and Fr **8** towards BAZ2A. Indeed, upon titration low intensity shifts were observed so they were discarded from the analysis because considered either non genuine or too weak binders. Details of the overlay of the spectra are reported in figure 3.13.

Fr **14** showed CSPs only against the PHD of BAZ2A, so was exclusively titrated against this protein. Analysis of the K_D values highlighted a first difference between the two proteins: values are higher for BAZ2A PHD than BAZ2B, implying a lower affinity for BAZ2A (Table 3.5).

At this stage, very little information is gained about the binding mode of these fragments but CSPs allowed to formulate hypothesis on the potential region of binding (figure 3.6 A-I).

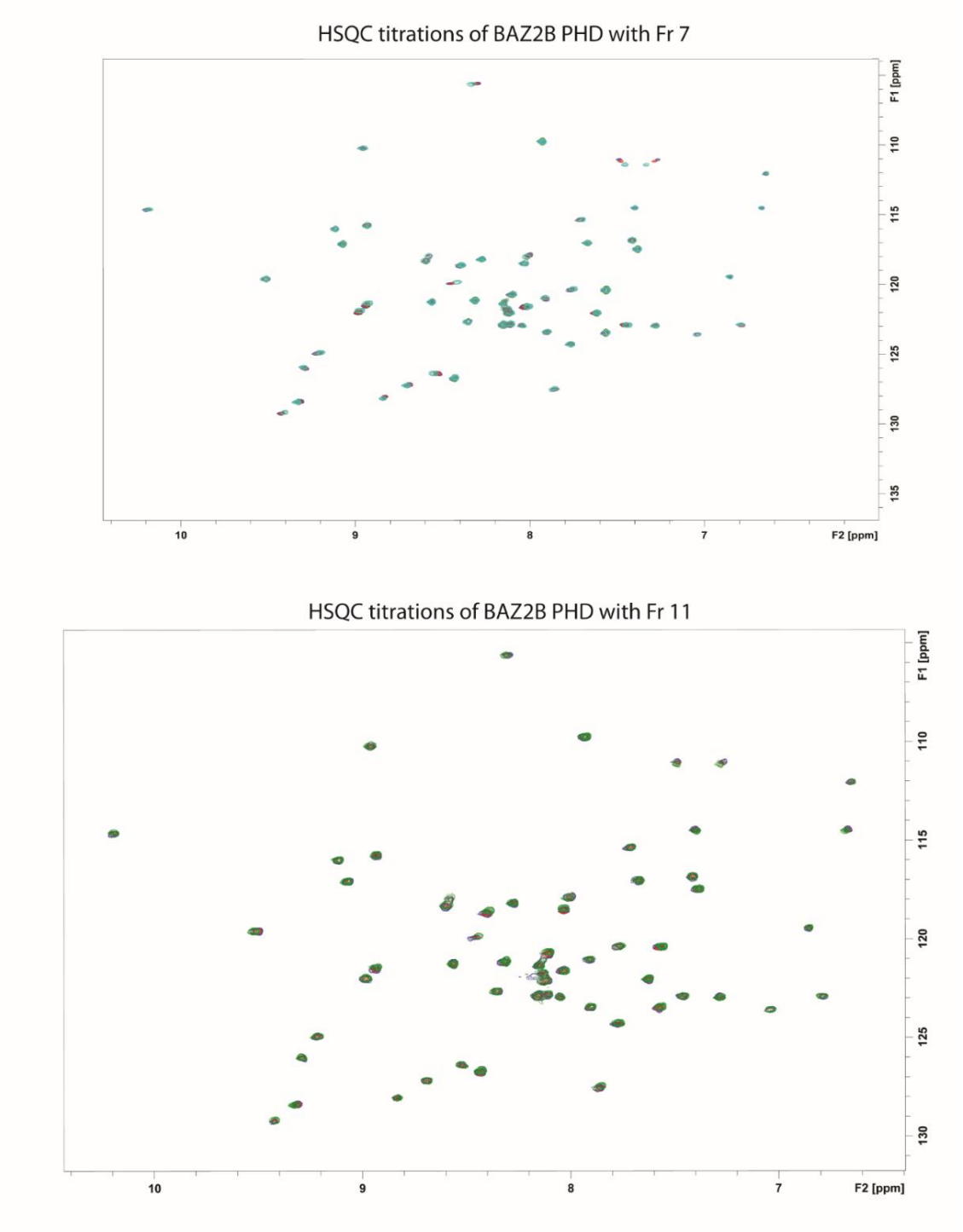


Figure 3.13. HSQC spectra.

Spectra reporting the HSQC titration of PHD of BAZ2B with Fr **7** (upper panel) and Fr **11** (lower panel). Meaningful shifts are observable only after high fragment concentration above 5 mM.

3.6.2 Surface Plasmon Resonance (SPR)

Another orthogonal technique was chosen with the intent to validate the binding affinities of this pool of selected fragments and compare with values obtained by NMR. Considering the number of fragments to evaluate, it was decided to use surface plasmon resonance (SPR). Kinetic and equilibrium binding parameters can be determined by SPR using a minimal amount of both protein and fragments. In the cases of genuine reversible binding ligands, K_D values can be obtained and are found to correlate quite well with K_D determined by other methods

3.6.2.1 Assay development: choice of the immobilization method

The first step was to set up the conditions of work for the assay and optimize the choice of immobilization of the protein.

There are different ways to immobilize a protein onto the surface of a chip: non covalent methods with high affinity interactions as, for example biotin/streptavidin, or covalent binding via free amine or thiol coupling. One of the most popular way of immobilization is the use of the streptavidin/biotin affinity, requiring as further step only the biotinylation of the protein. The risk associated with the choice of biotinylation is the generation of heterogeneous product with random orientation of the protein and consequently impaired function. Indeed, the main concern about the biotinylation of our PHDs is that these proteins are very small domains (~60 aa) containing only three or five Lys located in proximity of the two potential binding pockets. In this way, upon biotinylation and binding of the protein to the streptavidin coated surface, the PHD pockets could results as partially blocked from steric hindrance and with restricted access for the fragments. In addition, to support this theory there were results from previous fragment screening with BLI, using biotinylated protein, that did not lead to any interesting results for hits later confirmed by protein-observed NMR.

To obviate these issues, I decided to adopt an alternative strategy immobilizing the protein to the surface of the chip using the 6His affinity tag and stabilizing the binding through a capture coupling [155]. The aim of this approach is to expose both pockets, to control the steric orientation and to warrant a stable surface with higher binding

capacity. The 6His tagged constructs of PHD of BAZ2A and BAZ2B were found not soluble. Suitable for our purposes was, instead, the usage of the soluble construct of 6His-SUMO-PHD, that is purified prior to protease cleavage to release the tag (see Method 6.1.6). The main concern arising from the usage of this tag is that SUMO (MW ~12 KDa) may itself interact with small molecules. In order to account for this potential interactions and to compensate for any matrix effect due to interactions of the fragments with the surface of the chip, a reference cell was set up, containing the 6His-SUMO tag alone to be subtracted to the actual data obtained during the data experiment (figure 3.14).

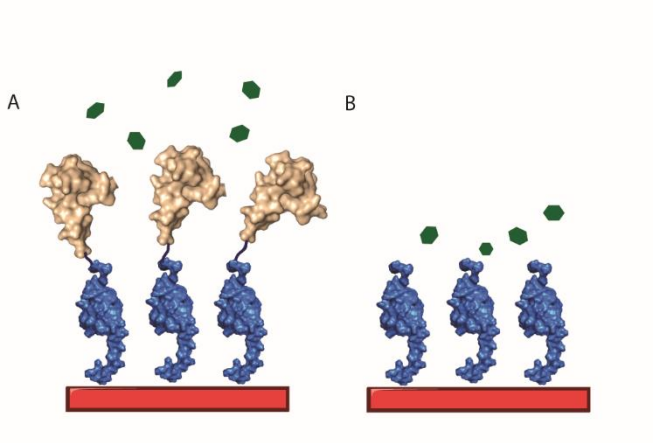


Figure 3.14. Schematic representation of the protein immobilization onto the sensor chip.

(A) Representation of the surface of the sample cell with 6His-SUMO-PHD protein immobilized onto the chip. In blue 6His-SUMO and in wheat the PHD. (B) Representation of the reference cell with the only 6His-SUMO protein immobilized. In green is depicted the analyte.

3.6.2.2 Assay set up

Proteins were immobilized using a Ni-NTA chip. I mainly followed the protocol reported in the paper from Kimple *et al*. [155]. The surface of the chip, constituted of carboxymethylated dextran pre-immobilized with nitrilotriacetic acid (NTA), was first regenerated with an EDTA solution to remove all potential contaminants, subsequently, a Ni^{2+} solution was injected to bind the surface. The surface was then activated with a coupling solution EDC/NHS and then the tagged protein was allowed to chelate the

metal ions. Ethanolamine was injected in the final step to block uncoupled primary amines on the surface. The capture coupling has the further purpose to stabilize the surface avoiding loss of protein over the course of the experiment.

The sensorgram obtained for the immobilization process of BAZ2A/B PHD is reported in Figure 3.15.

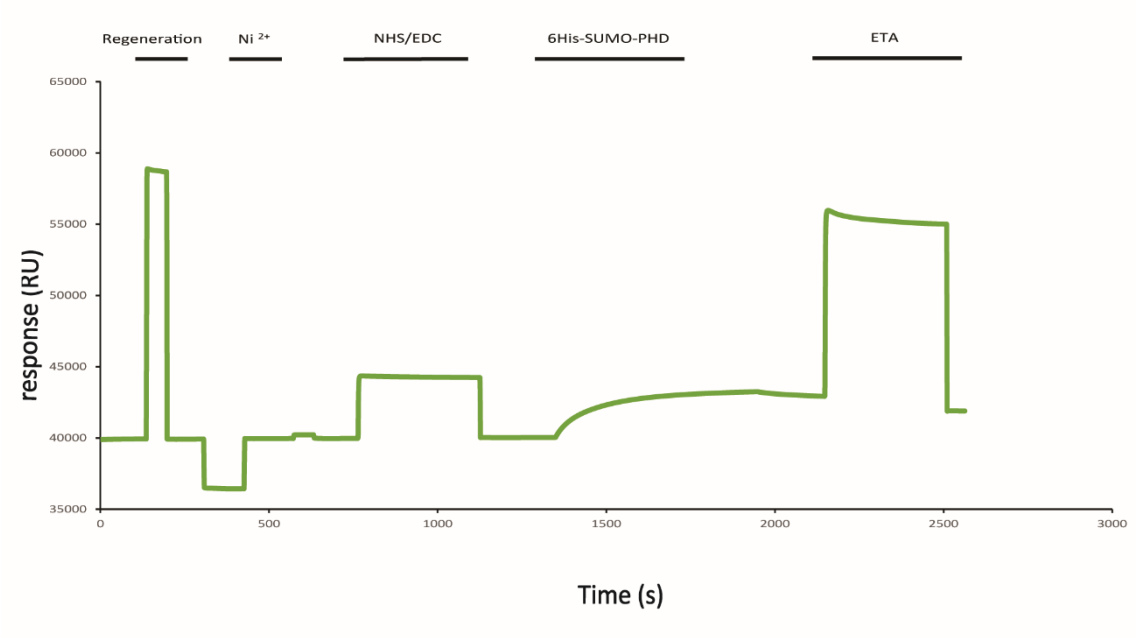


Figure 3.15. Sensogram showing the different steps for protein immobilization.

The sensor chip was first regenerated, and subsequently allowed to bind Ni^{2+} at the surface. The surface was then activated with the coupling solution EDC/NHS and then it was allowed the tagged protein to chelate the metal ions. Ethanolamine was injected in the final step to block uncoupled primary amines.

3.6.2.3 Analysis

Prior to the testing fragments, as positive control to guarantee that the immobilized protein was actually active, the H3 10-mer peptide was used [46]. As negative control was used one of the fragments, from the *in silico* screening (Fr **1**) that did not show any shift by HSQC NMR. H3 10-mer peptide and Fr **1** were both tested in three different concentrations in dilution 1:2 starting from 500 µM (10~fold K_D [1]). Responses detected are reported in the plot in figure 3.16 and are the difference between sample cell minus reference cell and further subtraction of the nearest blank. Peptide titration was repeated at the end of the assay to check the stability of the response over the time.

For each point the relative R_{max} was extrapolated. R_{max} corresponds to the maximum binding capacity, in RU, of the protein bound on the surface. R_{max} was calculated using the following formula, assuming that the protein is binding to one molecule of analyte (peptide or fragment):

$$Rmax = \frac{Rbaseline * MWanalyte}{MWprotein}$$

where the $R_{baseline}$ is representative of the amount of PHD present on the surface of the chip and is calculated as the difference between the baseline of the surface containing the 6His-SUMO-PHD and the reference cell containing 6His-SUMO protein alone; $MW_{anlayte}$ is the molecular weight of the peptide or fragment and $MW_{protein}$ is the molecular weight of the PHD of BAZ2A/B.

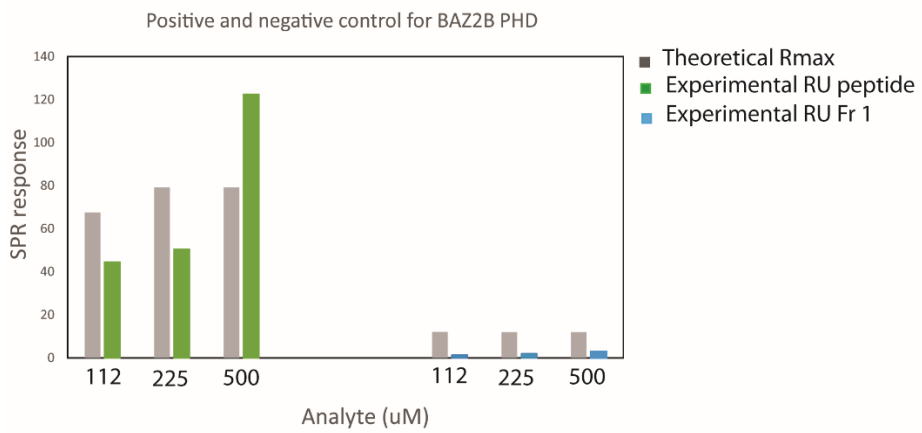


Figure 3.16. SPR response to H3 10-mer peptide and Fr 1.

Plot of the SPR response detected for H3 10-mer peptide (green) and negative control (Fr **1**, blue) towards BAZ2B PHD. Three concentrations tested are reported in µM on the x axis. In grey, for comparison, are reported the relative R_{max}.

As shown in Figure 3.16, for 500 µM peptide the experimental RU at the beginning of the titration is slightly above the R_{max} while the other two points are within the admitted range. This effect could be explained by dimerization of the peptide at high concentration and in presence of DMSO. The negative control showed low response as expected.

Subsequently, both proteins were screened with those fragments previously validated by HSQC: Fr **3**, Fr **5**, Fr **7**, Fr **8**, Fr **11**, Fr **18** and Fr **19** reported in table 3.3. Six different concentrations were injected, in a dilution 1:2 starting from 900 µM, for each fragment. Higher concentration of 900 µM was avoided in order to reduce mass transport effects due to a high concentration.

Fr **17** was not included due to availability issues at the time of the experiment.

The first observation from the analysis of the sensorgram was the low response detected for all the fragments, that does not allow a proper analysis. Responses detected were in the same range of the negative control. In figure 3.17 were plotted the RU/Rmax (%) for each concentration of each compound. This plot is informative of the fraction of maximal activity observed for each compound.

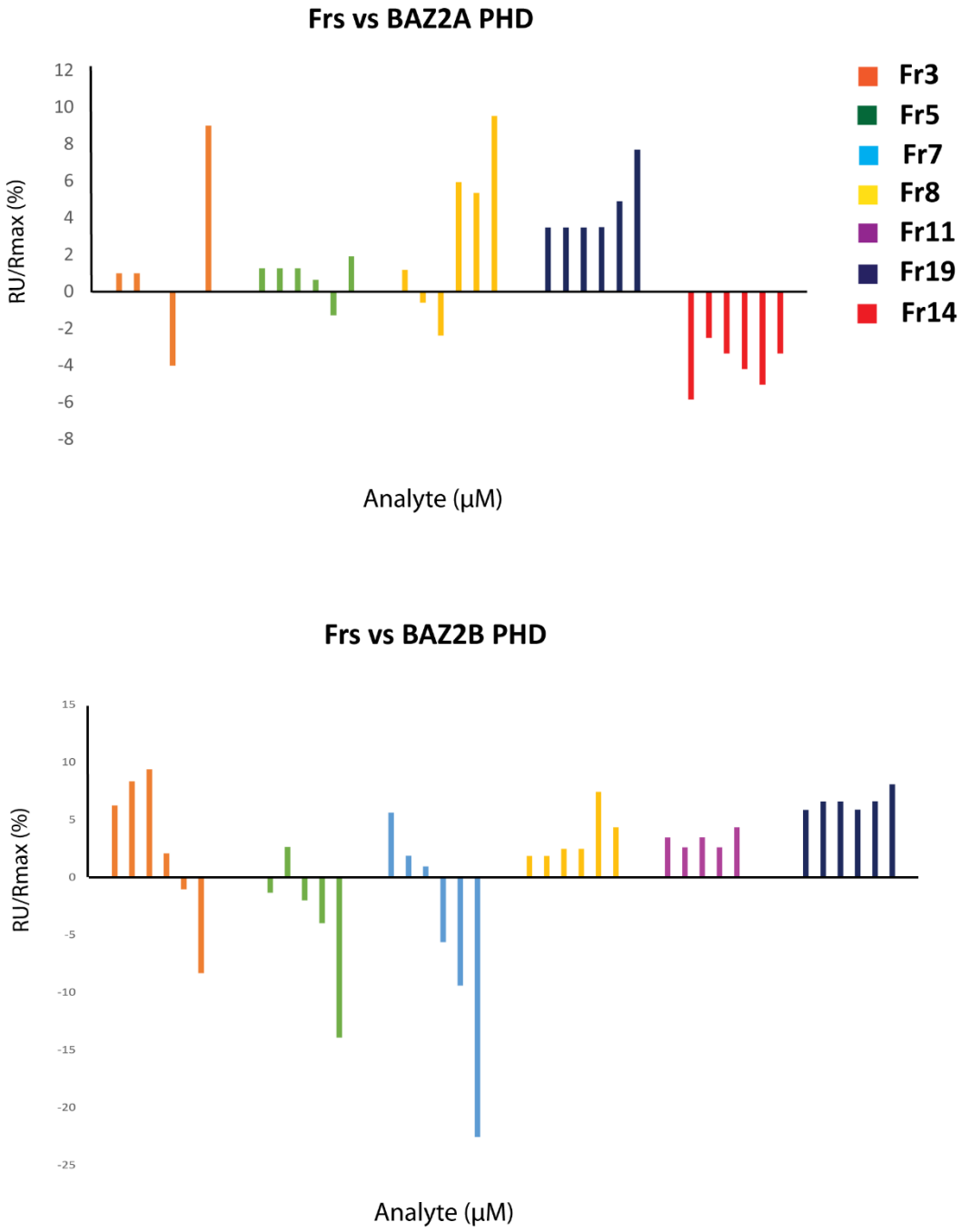


Figure 3.17. SPR responses upon fragment injections.

Histograms of the response detected by SPR for each fragment at different concentrations. In the upper panel the response versus BAZ2A PHD and in the lower panel the responses versus BAZ2B PHD. Fragments are grouped by colour as reported in the legend.

The responses detected were very low in RU but for some fragments it was still possible to observe a trend of increasing response at increasing concentrations. This is the case, for example, of Fr **19** towards both BAZ2A/B PHD, however the maximum active fraction

was lower than 10% of the expected one. Also Fr **3** in BAZ2B showed a concentration dependent response for the first three injections.

An interesting case was Fr **18** that for both proteins showed negative responses at high fragment concentration. Responses relative to Fr **18** were taken out of the analysis since towards both proteins were observed high negative response that would have masked responses of the other fragments. This event could be caused from the bulk-effect, due to the presence of precipitated compound flowing through the cells. The refractive index is dependent from the bulk properties of the buffer and the presence of small particulate or too high density of the solution could generate a negative response. Despite the several studies about SPR, there is no available literature able to explain this phenomenon.

The low responses showed in the sensograms seem to suggest different potential scenarios:

1) the interaction of the fragments with the PHD proteins is comparable to the interaction with the protein in reference cell;

2) the fragments either do not bind at all, or bind too weakly to the PHD domain to generate a measurable signal

3) the similarity in responses between data and reference cell could be determined from the partial degradation of the tagged protein 6His-SUMO-PHD into 6His-SUMO, still bound onto the surface, and PHD alone. In this situation the surface of data and reference cell would be exactly the same justifying the zero response. This last assumption is further supported by mass spectrometry analysis for BAZ2A protein that detected three species in solution: 6His-SUMO-PHD, the 6His-SUMO and PHD alone (figure A.6 in Appendix), suggesting instability of the protein and degradation over the time.

Another phenomenon to consider is the binding of some of the fragments to the Ni^{2+} on the surface of the chip. Indeed, this interaction may generate interferences which could lead to unreliable response in the assay. In particular, amongst the pool of fragments tested, Fr **8** (Table 3.1), which carries the sulphonamide group, could be capable of metal chelation on the surface of the chip and interference with the assay.

3.6.2.4 Limitations of the assay and alternatives

These data taken in account together suggest a main limitation of the usage of the SPR assay to detect fragment binding. It is possible that fragments under investigation bind too weakly to be adequately detected in the assay. In addition, the construct required for the immobilization of the protein (6His-SUMO-PHD) showed to be unstable and thus could be unsuitable for detection of reliable responses. Therefore, another strategy should be addressed in future. A potential alternative would be the usage of a tag that allow a stable immobilization, simultaneously warranting the right orientation of the protein, exposing both pockets. The AviTag would be an attractive choice. AviTag is a sequence of 15 amino acids (GLNDIFEAQKIEWHE) that can be enzymatically biotinylated on the single Lys residue present in the tag sequence using the enzyme biotin ligase (BirA), in presence of biotin and ATP [156, 157]. This peptidic sequence can be inserted by inverse PCR mutagenesis [157]. The advantage is the selective biotinylation of the Lys on the AviTag, preventing derivatization of lysine residues that could affect protein activity. In this way it would be possible to immobilize the protein on the surface of the chip in a controlled manner, exposing all the potential binding sites, as expected. The more general advantage of this methodology is that the usage of this tag could be applied to other high-throughput screening that requires a stable and oriented immobilization of the protein.

3.7 ALPHALISA COMPETITION ASSAY

One of the limitations of the chemical shift mapping is that it may be difficult to distinguish whether it arises from specific or non-specific binding since broadening and change in intensity of the peaks could be caused from either non- specific binding or intermediate exchange. In this context, application of an orthogonal functional technique, such as competition AlphaLISA assay, could provide additional information useful to elucidate the binding mode of the small molecules.

In order to test if the identified hits were able to bind the PHDs BAZ2A/B in the histone pocket, I developed a competition assay using the AlphaLisa bead-based technology.

3.7.1 Assay development

The assay was set up using the streptavidin donor beads and the anti-FLAG acceptor beads. Proteins were biotinylated in order to bind the donor beads. For this assay, the usage of biotinylated protein does not constitute a limitation since the signal observed is considered to arise from the fraction of active protein on the bead surface. Even if this fraction is low, it must be sufficient to generate a signal that will be still appropriate to perform competition studying by fragments.

The peptide H3 10-mer AA mutant [46] was chosen for the assay and synthetized with the additional Flag epitope at the C-terminus (ARTAATARKS-DYKDDDDK), in order to be recognized by the anti-FLAG acceptor beads. The binding of the FLAG tagged peptide was confirmed for both PHDs BAZ2A/B by ITC (Table 3.6 and figure 3.18) and the obtained K_D correlated well to those obtained for the H3 10-mer AA mutant (BAZ2A K_D ~12 µM and BAZ2B K_D ~ 2.6 µM)[46].

Protein	K_D (µM)	N	ΔH (kcal/mol)	TΔS (kcal/mol)	ΔG (kcal/mol)
BAZ2A PHD	5.9 ± 0.08	1.27 ± 0.002	-7.16 ± 0.01	-0.03 ± 0.02	-7.13 ± 0.01
BAZ2B PHD	1.7 ± 0.05	1.29 ± 0.002	-8.3 ± 0.01	-0.5 ± 0.02	-7.8 ± 0.02

Table 3.6. Summary of thermodynamic binding data for FLAG peptide.

ITC-based binding parameters for complex formation between FLAG peptide and BAZ2A/B PHD.
Error values reported on K_D, N, and ΔH are generated by the Origin program and reflect the quality of the fit between the nonlinear least-squares curve and the experimental data. Errors reported on TΔS and ΔG were propagated from the errors of K_D and ΔH.

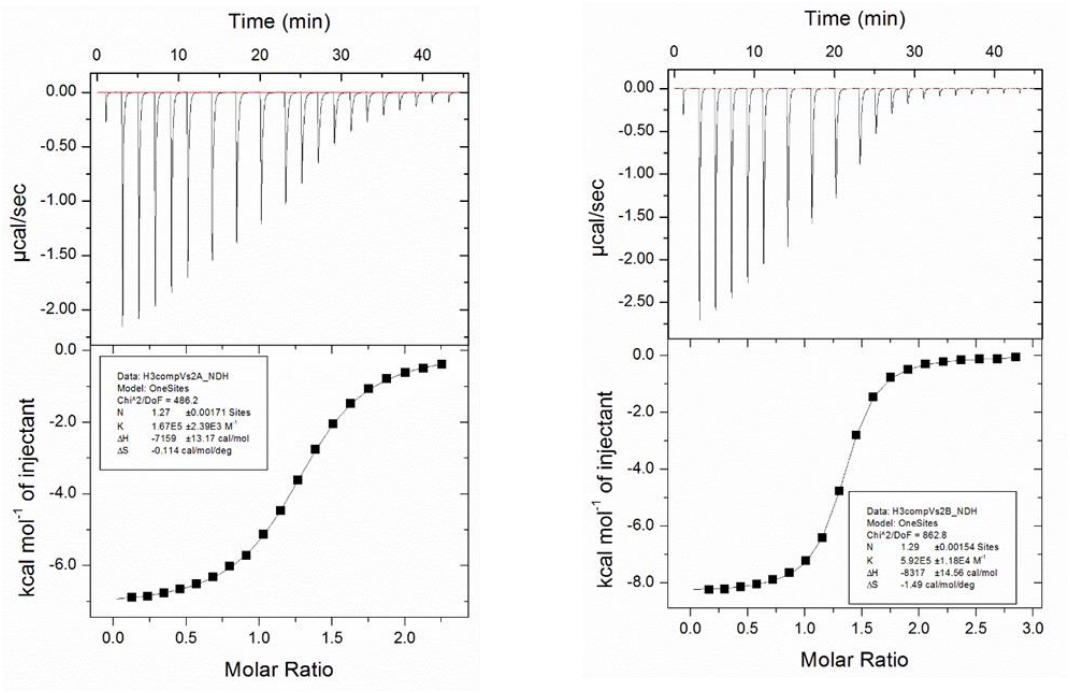


Figure 3.18. Raw ITC experiments with the FLAG-peptide.

Raw ITC binding curves of H3 FLAG peptide into PHD of BAZ2A (left) and BAZ2B (right) are reported in the upper panel. The integrated ΔH (kcal/mol) values are plotted versus the peptide/protein molar ratio and shown in the lower panel.

Prior to performing the assay, protein and peptide concentrations were optimized. Using 5 µg/mL of both beads, different protein and peptide concentrations were screened in order to get the optimal signal over the background range but to avoid the hook effect that can occur when in solution there is an excessive amount of protein or peptide to saturate all the binding sites, inhibiting the interaction between the beads. A first protein/peptide screening did not generate any signal of interaction between peptide and protein. This result suggested that something within the assay could prevent the binding between protein and peptide, as for example the length of the peptide. Indeed, the FLAG tag, immediately adjacent to the H3 epitope could have impeded the correct structural organization of the peptide, needed for the binding with the protein. In order to overcome this issue, a longer peptide was synthesized carrying an extra flexible linker (TGGSGGSG) between the N terminus and the FLAG epitope (ARTAATARKS-TGGSGGSG-DYKDDDDK, namely AlphaLISA peptide). Furthermore, DTT was removed from the assay since it reduces the disulfide bonds on the M2 epitope of

the anti-FLAG beads, needed to recognize the FLAG sequence. Using the long form peptide and in absence of DTT, signal was now pleasingly observed (Figure 3.19). A peptide concentration of 160 nM was sufficient to produce maximal response. A 10 nM concentration was chosen for both proteins since it gave reasonable signal without saturating the system, simultaneously guaranteeing a low usage of protein.

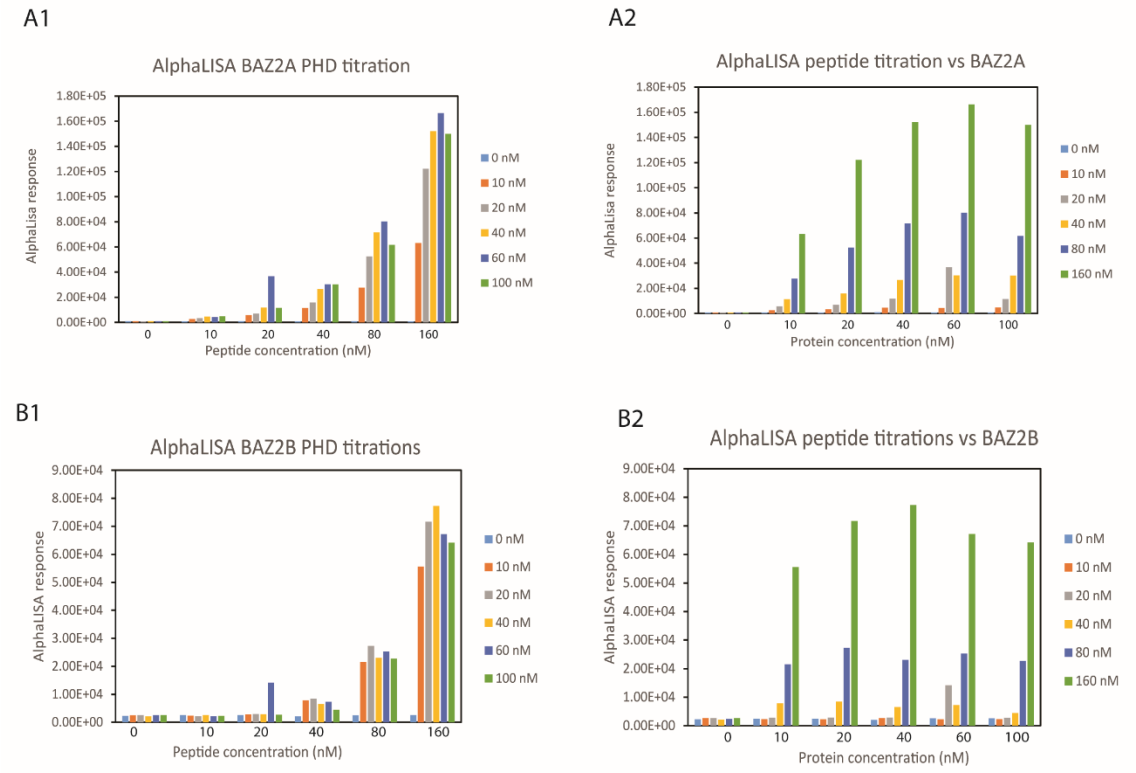


Figure 3.19. Determination of saturation points of acceptor (protein) and donor (peptide) in AlphaLisa assay.

Histograms reporting AlphaLISA response versus peptide/protein concentrations. The hook effect is observed increasing the protein concentration.

(A1) and (B1): BAZ2A/B PHD titrations with six different peptide concentrations.

(A2) and (B2): peptide titrations with six point protein concentrations of BAZ2A PHD and BAZ2B PHD, respectively.

AlphaLisa displacement assay was then set up using 10 nM of protein and 160 nM of peptide in 100 mM HEPES buffer supplemented with 0.02 % CHAPS and 0.1% BSA to minimize non-specific binding. No salt was added in order to minimize extra interferences with the assay. Buffer concentration was kept high to avoid pH changes at high fragment concentrations.

Validity of the assay was assessed for each protein using the S/B and Z' factor [158], calculated as described in chapter 6.4.4. Table 3.7 summarizes the results.

Protein	S/B	Z-factor
BAZ2A PHD	~72	0.79
BAZ2B PHD	~22	0.43

Table 3.7. S/B and Z-factor for BAZ2A and BAZ2B PHD AlphaLISA assay.

Signal to background (S/B) is large for both proteins. Z-factor is good for BAZ2A since over 0.5, which is considered the optimal threshold [158]. In BAZ2B the Z-factor is lower than 0.5, nevertheless can be considered an acceptable value to state reasonable quality of the assay since the range of points used for the analysis is low (four points) and the deviation from the threshold is still minimum.

3.7.2 Competition assay with H3 10-mer peptide

Prior to performing AlphaLisa competition assay with the fragment set, a positive control experiment was performed. The H3 10-mer wild-type peptide was used as competitor. Eight different concentrations of peptide (in serial dilutions 1:5) were tested and IC_{50} was obtained from the dose-response curves (figure 3.20 B-C) with the relative 95% confidence interval and plotted as $LogIC_{50}$ in figure 3.20 A. As negative control was used one of the fragments, Fr **1**, that did not show any shift in HSQC validation. Fr **1** did not generate any response in AlphaLISA, as expected.

3.7.3 Competition assay with fragments

Initially, only three concentrations of fragments were tested in order to explore the competition response with this assay set up.

Subsequently, a proper analysis dose-response was performed. Both PHDs were titrated against eight different concentrations of each fragment (dilution series 1:5). IC_{50} values

were extrapolated to assess whether fitting of the dose-response curve was possible. Fragments were tested according to their availability in the lab (Fr **3**, **5**, **7**, **8**, **14** and **18**) at the time of the experiment and the LogIC_{50} values of those fragments that gave a dose-response curve was plotted in Figure 3.20 D-E.

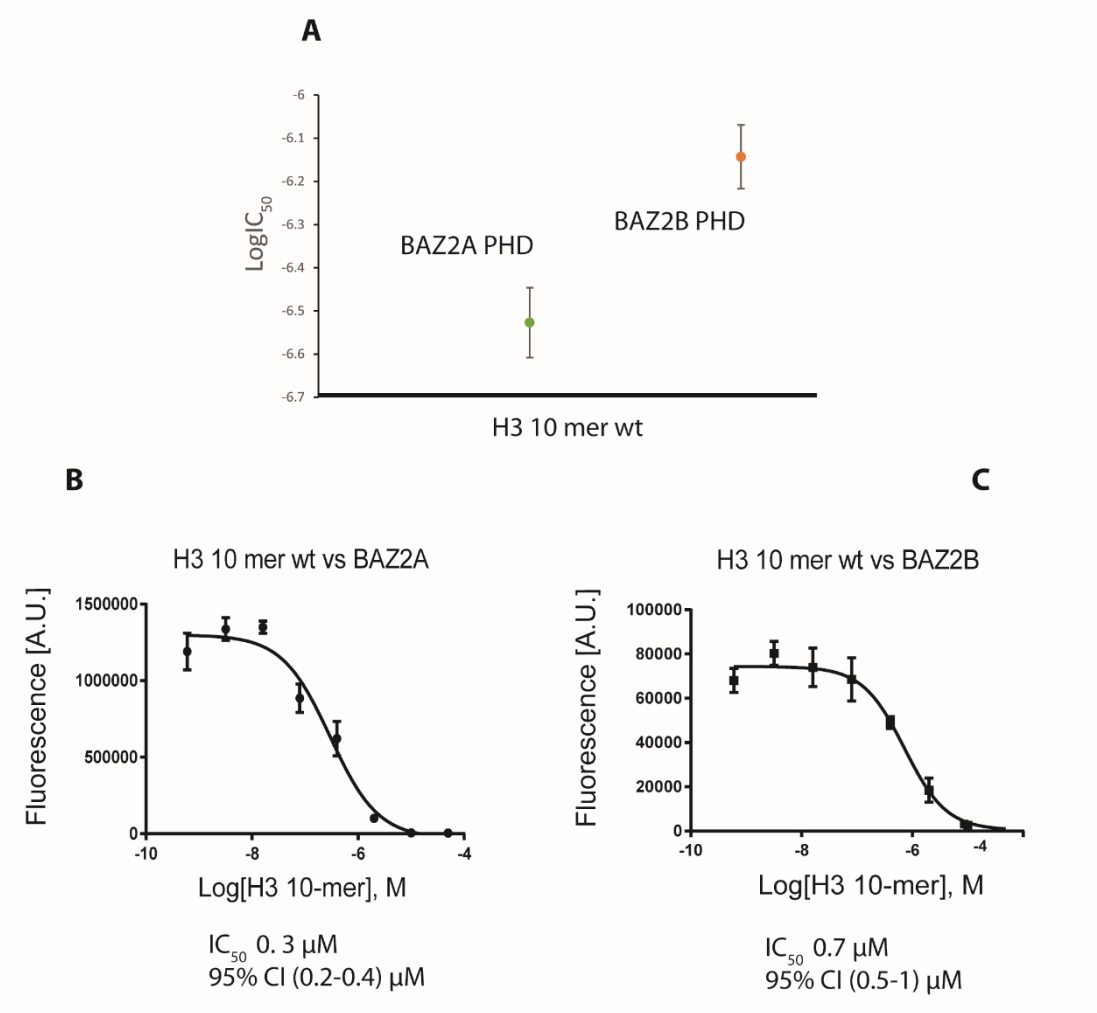


Figure 3.20 A-C. Results from AlphaLISA competition assay.

(A) Log(IC_{50}) were plotted against H3 10-mer wt peptide. In green the response detected towards BAZ2A PHD and in orange towards BAZ2B PHD. (B) and (C) dose-response curve obtained respectively for BAZ2A and BAZ2B PHD plus the respective IC_{50} extrapolated from the curves.

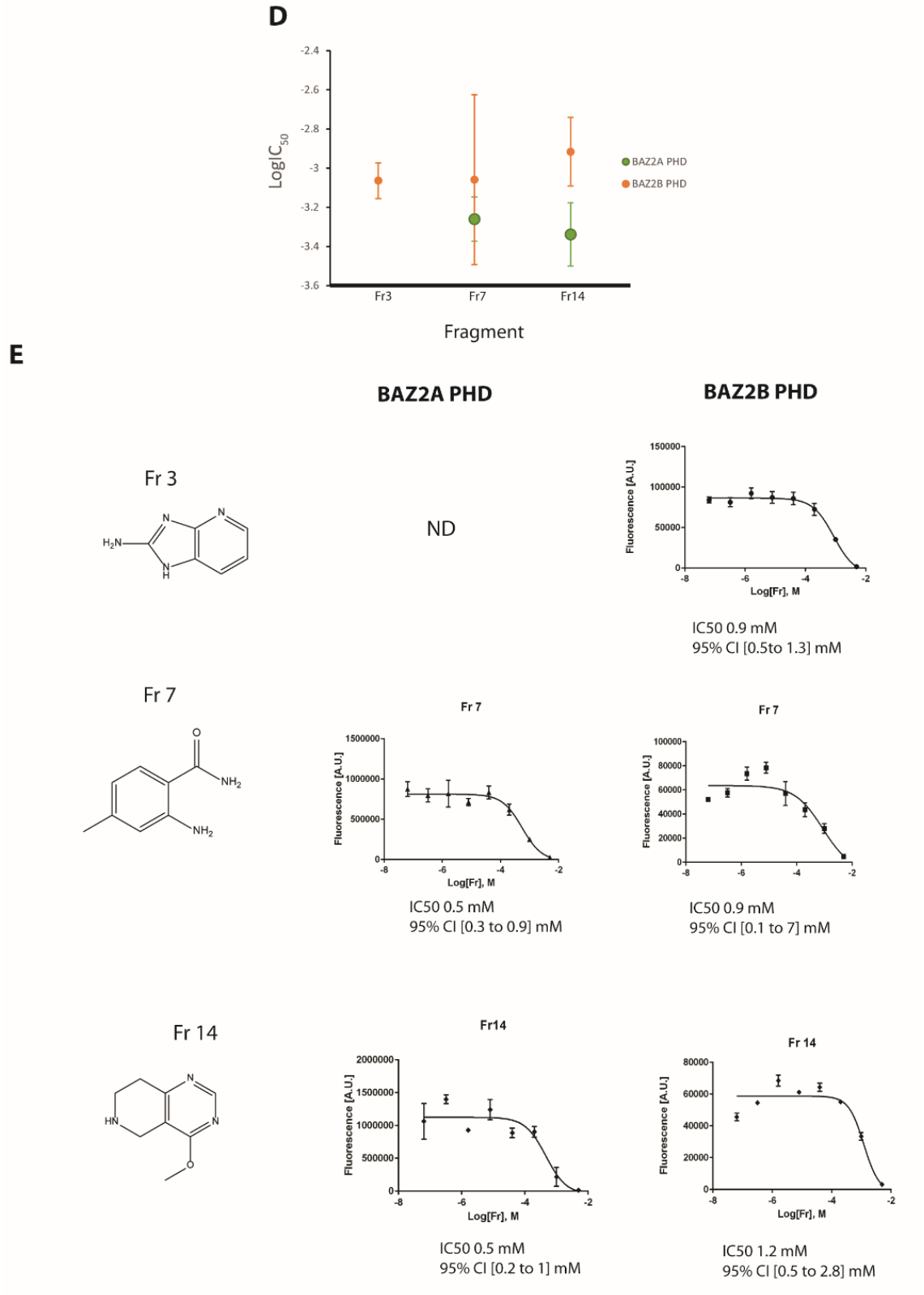


Figure 3.20.D-E. Results from AlphaLISA competition assay.

In panel D, Log(IC_{50}) were plotted against each fragment tested. In green the response detected towards BAZ2A PHD and in orange towards BAZ2B PHD. In panel E, raw dose-response curves. ND=not detected.

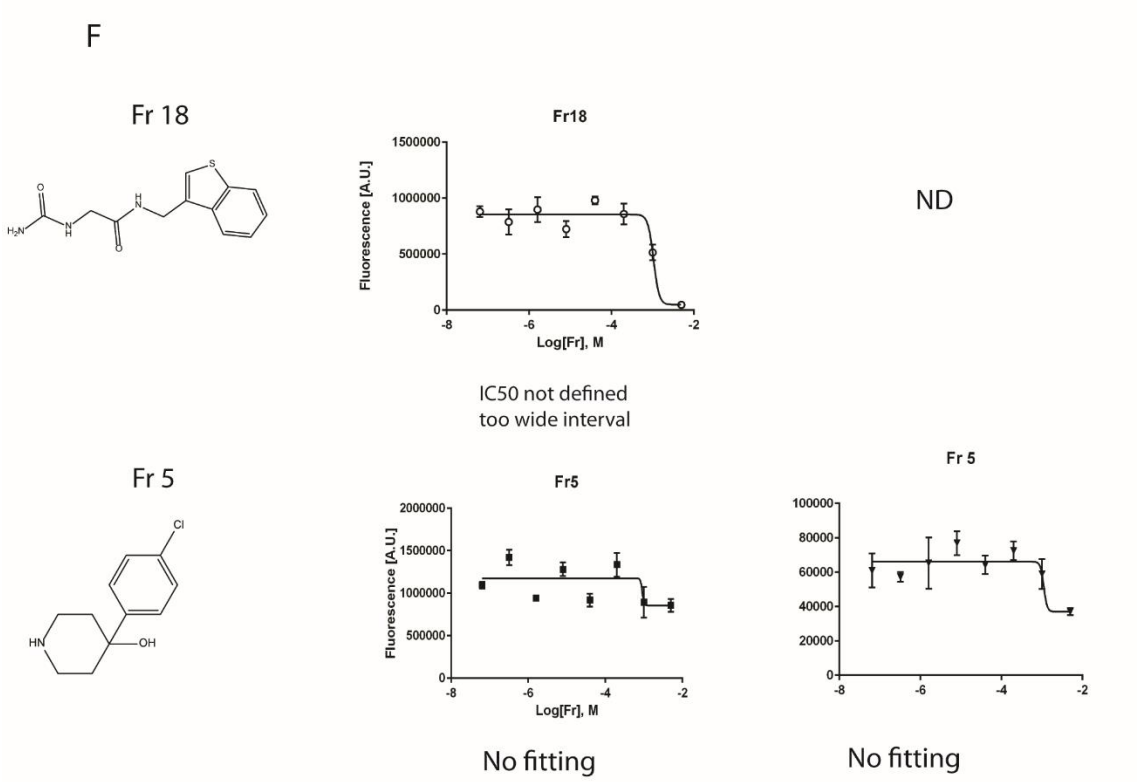


Figure 3.20.F. Results from AlphaLISA competition assay.

In panel F, raw dose-response curves of Fr **18** and Fr **5**. ND=not detected.

Reasonable dose-response curves were detected for Fr **3**, Fr **7** and Fr **14** (Figure 3.20 D-E). Unexpectedly, Fr **7** showed to interfere with the histone in both proteins, despite from a first analysis based on the CSPs was identified as potential binder only for BAZ2B. Similarly Fr **14**, that was supposed to interact only with PHD of BAZ2A, shows displacement also for the PHD of BAz2B although with a higher IC_{50}. A likely explanation of the phenomena could be either a non-genuine effect due to the high concentration of compound in the assay or an allosteric effect on the histone pocket.

Fr **5** did not seem to affect the interaction for both proteins at any of the concentrations tested. This might be indicative of the fact that is either non-specific binder or is binding a region of the protein, far from the histone pocket or its affinity is too low to generate any effect (figure 3.20 F).

Fr **18** generated a reasonable response only for BAZ2A PHD and at very high fragment concentration (over 1 mM) but the data could not be fitted to extrapolate an IC_{50} (figure 3.20 F). This may be representative of the fact that the fragment does not bind the

histone pocket (as hypothesized from CSPs) but the signal is decreased at high concentration of compound, potentially due to mass effect.

3.8 SOAKING EXPERIMENTS AND X-RAY STRUCTURAL STUDIES

3.8.1 Inspection of crystal packing

Structures of the *apo* form of BAZ2A and BAZ2B PHD were previously solved from Dr. Cynthia Tallant, a former postdoc in the lab [45]. The proteins crystallize in different space groups and exhibit different crystal packings, despite crystallizing in the same buffer conditions (2.2M Na/K phosphate buffer pH 8.5) (figure 3.21). These differences are summarized in Table 3.8.

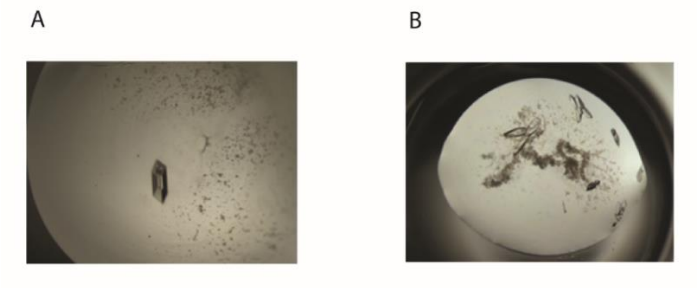


Figure 3.21. Picture of crystal of BAZ2A PHD (left) and BAZ2B PHD (right).

Crystals are obtained in 2.2 M Na/K phosphate buffer at pH 8.5 at room temperature.

Protein	Space group	N° protomers	Histone pocket	Second pocket
BAZ2A PHD	tetragonal $P4_32_12$	4	2 free 2 partially occluded	4 engaged in tight crystal packing
BAZ2B PHD	orthorombic $P2_12_12_1$	2	2 partially occluded	2 engaged in tight crystal packing

Table 3.8. Space group of crystallization for each protein and relative number of protomers per ASU.

Based on the observations reported in table above, I surmised that the two apo forms were not suitable soaking systems for the second pocket of the PHDs of both BAZ2A/B.

On the other hand, the histone pocket of BAZ2A was considered available to soaking because free from crystal contacts.

3.8.2 Solvent test for soaking

Soaking fragments into apo form crystals is one of the methods of choice to obtain ligand-protein complex owing to the ease of the process. Despite this, there are some limitations when using this approach and several factors to be considered. Indeed, crystals may be fragile and not tolerate soaking. Another limitation is represented by the solubility of the fragments in the crystallization conditions. It is often found that fragments are not soluble in the crystallization conditions and, consequently, some organic solvents are needed to increase their solubility but the addition of the solvent may destabilize the crystal packing.

To overcome this issue, the stability of both apo form crystals of PHD of BAZ2A/B was tested at different percentage of several solvents. In the first instance, I tested the stability in presence of DMSO, since is the most widely used organic solvent to dissolve small molecules. Unfortunately even low concentrations as 0.5% (v/v) of DMSO were not tolerated as the crystals immediately dissolved. A solvent screening was then performed testing the stability of the crystals in a solvent range between 2.5% and 20% (v/v). Crystals were incubated at different percentage for at least 1h before to be soaked in a higher percentage of solvent. Crystals were found stable in acetone, acetonitrile, 1,4-dioxane, ethanol and xylitol up to 20% (v/v) of solvent.

The diffraction pattern of crystals incubated overnight at 20% in the tolerated solvent was then assessed, using the appropriate cryoprotectant (20% glycerol). All of the above solvents resulted in good diffraction pattern with well-spread and not overlapped spots. This was in spite of the overall appearance and morphology of the crystals not being preserved after soaking in all cases.

3.8.3 Soaking fragments

The privileged system used for soaking was BAZ2A because of the presence of two accessible pockets in the ASU. However, also BAZ2B was soaked since small molecules

such as fragments can readily diffuse through solvent channels and bind to small pockets of the protein surface, even if partially occluded by crystal packing. It was therefore considered of interest at this stage to explore the ligandability of any potential sites on the protein surface.

Fr **3** and Fr **8** were soaked into crystals of both proteins even if they showed very nice shift only for BAZ2B PHD. The main issue for these fragments was their poor solubility in all the solvent tolerated from the crystals. Indeed, they were insoluble in dioxane, acetone and acetonitrile at the target soaking concentration of 100 mM. Because of this, the actual concentration used in the soaking could only be estimated and it spanned in a range between 10 mM and 50 mM. It was not possible to increase the amount of compound in the reservoir since when mixed with the crystallization buffer it generated crystallization of the components. Before mixing fragment and reservoir, to make a homogeneous suspension, the powder was vortexed for few seconds. Soaking was performed from 10 to 36 h, but despite the appearance of coloured crystals, no extra electron density was observed in any of the data set collected for these two fragments. For Fr **3** and Fr **8**, because of their low solubility, co-crystallization trials were also attempted but unfortunately to no avail.

Fr **7** and Fr **11** were found soluble up to 1 M in 1,4-dioxane, acetonitrile and acetone. Soaking was performed only with BAZ2B crystal since CSPs remarkable showed shifts only for this protein. For these fragments binding was expected in the second pocket that in the crystals is actually involved in crystals contact. If this assumption, based on HSQC data, was correct we would have expected to destroy the crystal packing after prolonged soaking or in alternative we might have found the ligand to bind in another region. Crystals soaked in a range between 25 mM-200 mM over 10/16 h were still showing a good diffraction pattern. However, the data sets collected did not show any extra electron density to fit the ligand. This may be indicative of a low binding affinity, in agreement with the HSQC data, that upon titration showed very little intensity of shifts (Figure 3.13).

Fr **14** was partially soluble in water at 1 M but once dissolved at 100-200 mM in the

soaking drops generated big crystals with a shape similar to the protein crystals. This phenomenon was not reduced using lower concentrations of fragments in the drop or changing solvents. BAZ2A crystals upon soaking with Fr **14** were found to colour in yellow, potentially suggesting fragment binding, but no electron-density was observed for the compound.

Fr **17** was soluble at 100 mM in water but crystals did not tolerate higher concentration than 5-10 mM. Also for this fragment no extra electron density was detected.

Fr **19** was soluble up to 1 M either in water or crystallization buffer and was then soaked in crystals of PHD of BAZ2A in a concentration range between 25 mM to 100 mM.

All the crystals, before to be flash-frozen, were cryo-protected in the same soaking solution supplemented with 20% glycerol.

3.8.4 Fragment 19

From the soaking of the Fr **19** within BAZ2A PHD crystals, one data set was collected at 1.9 Å resolution. Inspection of the electron density after data processing led to the identification of an extra density in the region of the histone pocket of chain B. Unexpectedly, no electron density were visualized in chain A and D (which exhibit exposed histone pocket), but only in B and C (where histone pockets are partially occluded by crystal contact). This could be in agreement with the fact that in the solvent exposed pockets, since the affinity of the fragment is too low, the solvent diffusion is too fast and does not lead to any interpretable electron density. Indeed, the two pockets of chain B and C are partially occluded and this would reduce the exchange rate leading to higher occupancy of the ligand in the pocket. Ligand Fr **19** was then modelled into the Fo-Fc map. Several refinement cycles were performed and the summary of the crystallographic data is reported in Table 3.11.

Analysis of the structure (Figure 3.22) allowed identification of important features of binding. The aromatic ring of the thiophene is not visible in the electron density suggesting that this portion of the molecule is not making stable interaction with the

protein. The alkyl ring is only partially visible, with density detected up to the two carbon atoms adjacent to the nitrogen. Omit map and polder map calculation [159] did not show any further density on this region of the molecule.

Binding is driven by the amino terminal group of the molecule that is involved in hydrogen bonds with the side chain of D1695 and with the carbonyl group of the backbone of L1693. Further electrostatic interactions are established with the backbone C=O of the G1716 (Figure 3.22 A). Hydrogen bonds are observed also between the carbonyl of the ligand and the backbone NH of the backbone of L1693 (Figure 3.22A).

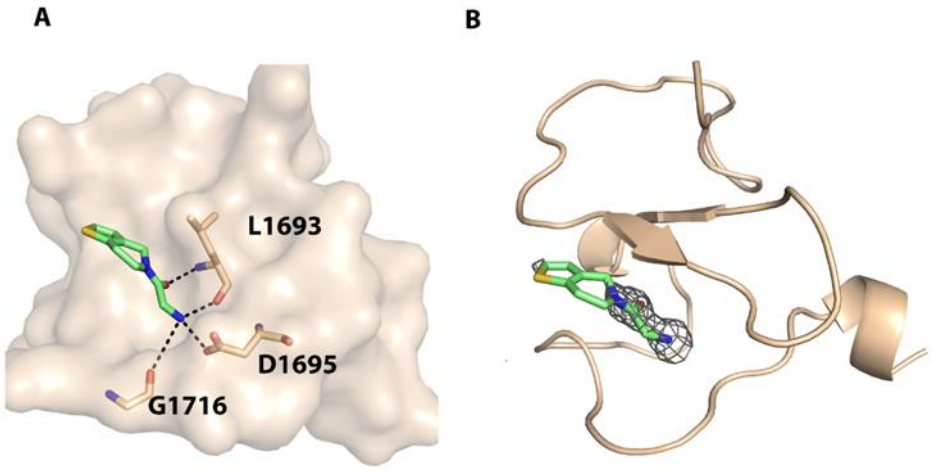


Figure 3.22. Crystal structure of the complex of BAZ2A PHD with Fr 19.

(A) Frontal view of the surface and cartoon representation (wheat) of BAZ2A PHD in complex with Fr **19** (cyan sticks). Residues of the protein interacting with the ligand are represented as stick and labelled in black. (B) Unbiased Fo-Fc electron density map is reported in grey and countered at 3σ.

3.9 FRAGMENT OPTIMIZATION

Analysis of the histone pocket of the fragment-bound structure of BAZ2A PHD with Fr **19**, showed how the particular conformation and structure of Fr **19** did not allow the thiophene ring to fulfil the hydrophobic region of the histone pocket (in yellow in figure 3.23) so this part of the molecule needed further optimization.

Molecular modelling studies were performed by Dr. Lucas Xavier with the aim to improve the binding mode of the fragment. The 2-amino-acetamide region was conserved and kept fixed in the conformation identified in the crystal structure (in red

in the chemical structure of Fig. 3.23) but flexibility of the molecule was improved substituting a tertiary amide with a secondary amide. Docking sought to identify the best group to fill the hydrophobic cavity (yellow surface in figure 3.23), for example a phenol or thiazole group (Table 3.9).

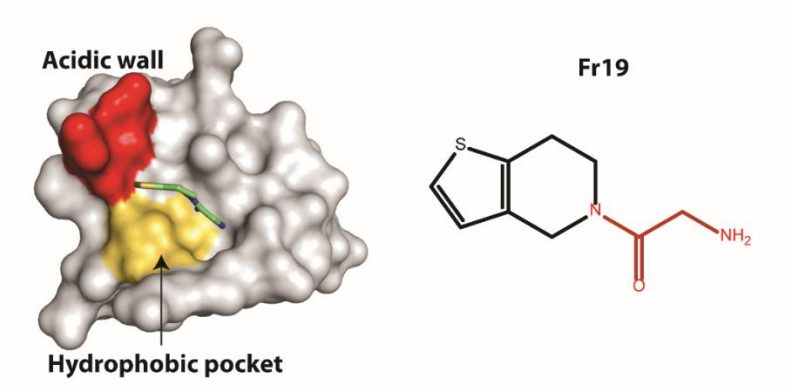


Figure 3.23. BAZ2A PHD in complex with Fr 19.

On the left, surface representation of BAZ2A PHD in complex with Fr **19** (green stick). In red the surface of the acidic wall and in yellow the surface of the hydrophobic pocket. On the right, chemical structure of Fr **19** showing in red the portion of the molecule which was conserved during the optimization.

In order to further stabilize the aromatic ring in the pocket we sought to catch hydrogen bonds with the side chain of the residue D1688 on the acidic wall (Fig. 3.23) introducing either a phenol or aniline group (Table 3.9).

Fr **20**-**23** were purchased and tested by biophysical techniques.

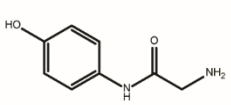
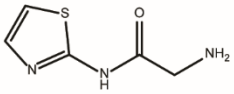
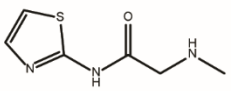
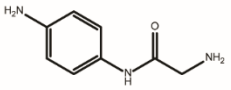
Fragment 20	Fragment 21	Fragment 22	Fragment 23
			

Table 3.9. Chemical structures of fragments obtained upon optimization of Fr 19.

3.9.1 HSQC and K_D determination

All four fragments were first validated by singleton HSQC. Then, systematic titrations were performed in order to estimate a K_D of binding. These optimized fragments were all soluble in water and NMR buffer, with the exception of Fr **20** soluble in water only at 100 mM. Titrations were performed up to 5 mM as for the previous fragments. Figure 3.24 A-B reports a summary of the extrapolated K_D for each protein and heat map obtained from CSPs. In red the strong shifts (above $\overline{\Delta\delta}$+ 2σ) and in orange the intermediate shifts (above $\overline{\Delta\delta}$+ σ).

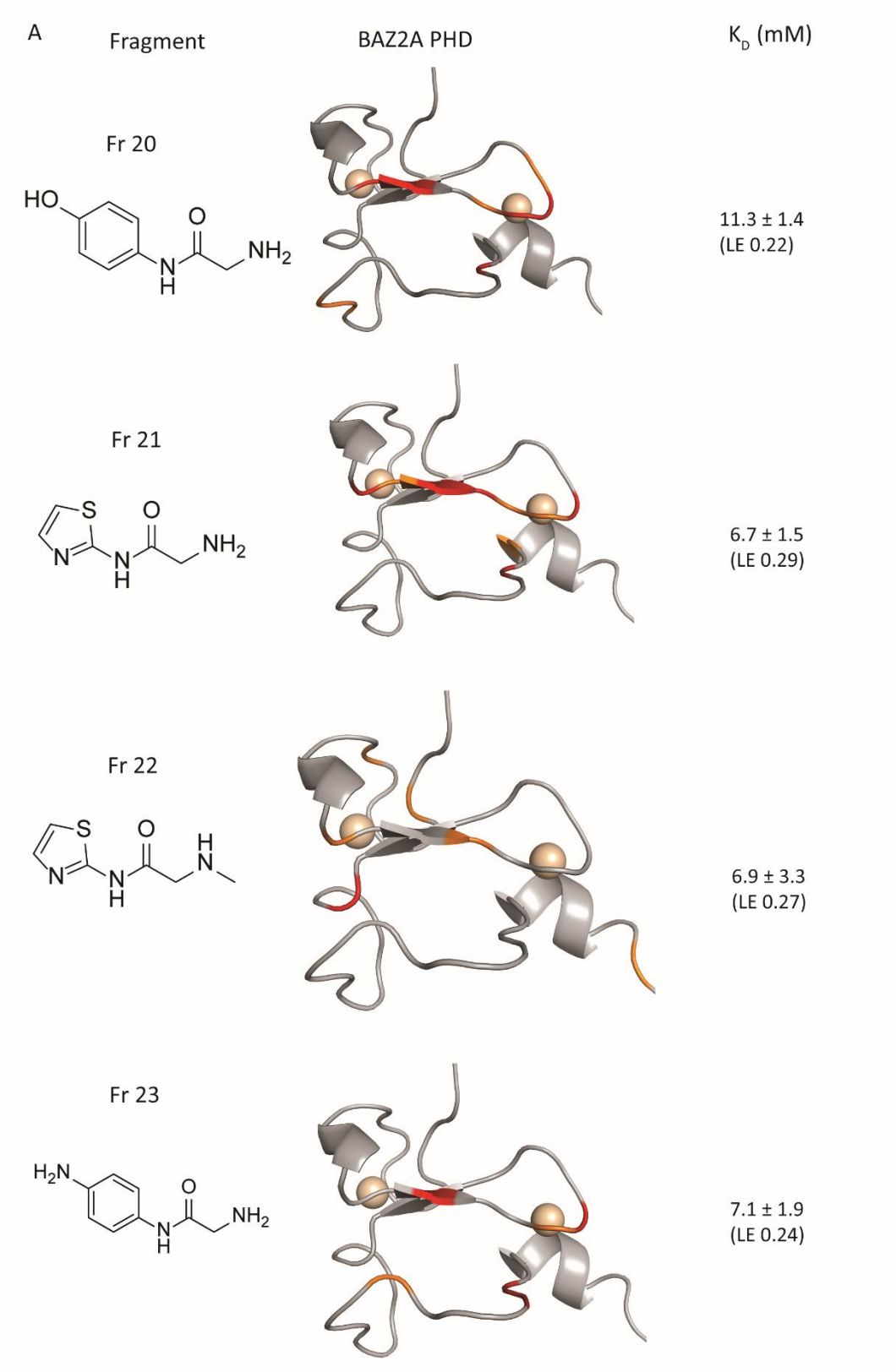


Figure 3.24.A. CSPs and K_D estimation for the binding of fragments analogue of FR 19 to BAZ2A PHD

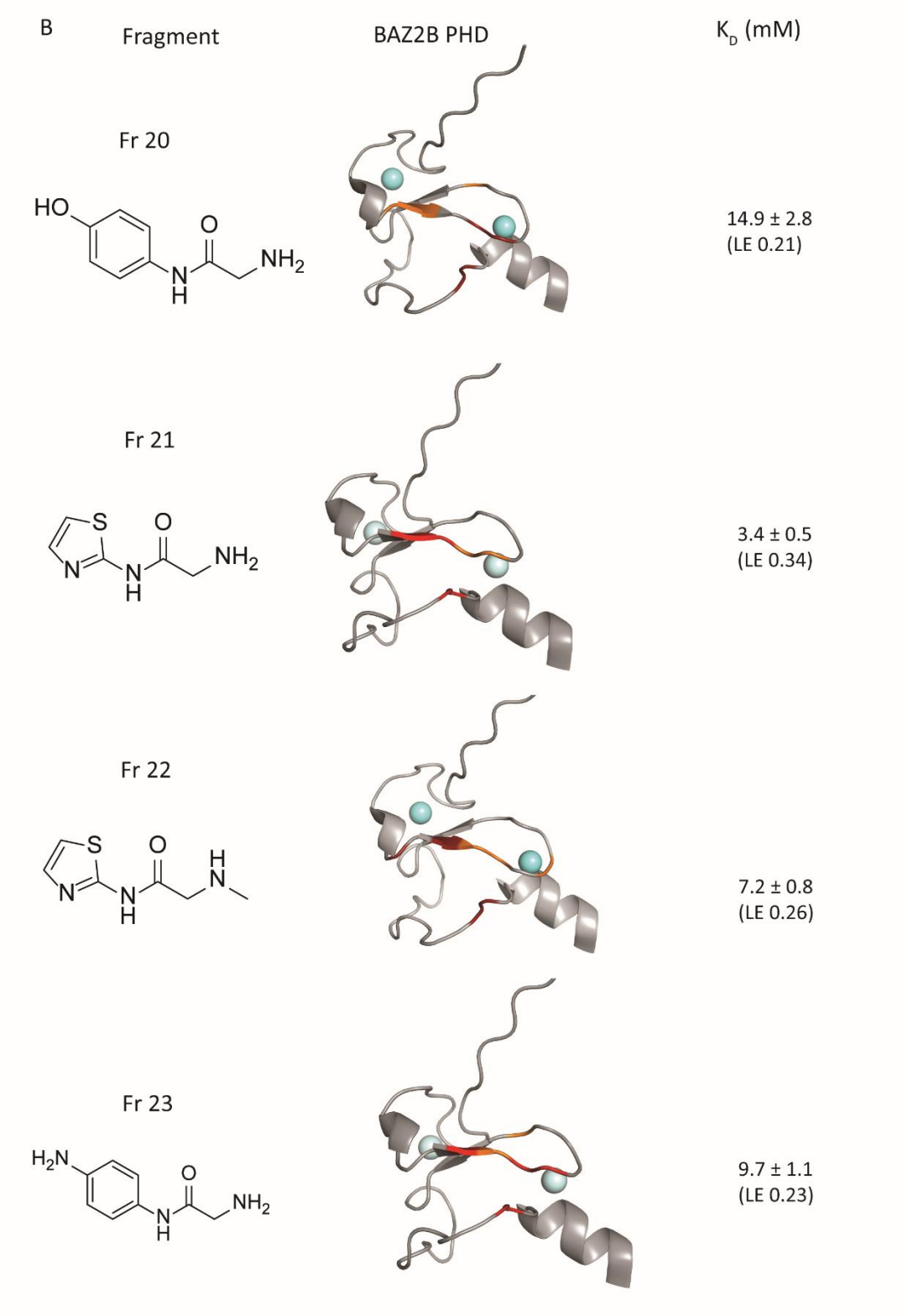


Figure 3.24.B. CSPs and K_D estimation for the binding of fragments analogue of FR 19 to BAZ2B PHD.

CSPs reported for the optimized fragments (left column) were mapped on BAZ2A (PDB: 4QF2, panel A) and BAZ2B (PDB: 4QF3, panel B) PHD structures.

Fr **20**, **21** and **23** shows clear concentration depending shifts without any effect on the intensity or shape of the peaks, in agreement with the expected fast exchange. Fr **22** showed a different behavior. A defined cluster of residues located in the histone pocket was observed in BAZ2B, as expected, and the intensities of shifts were comparable to those of the other fragments. In BAZ2A, the fragment showed less intense shifts involving residues randomly distributed over the protein surface. This behavior could point toward non-specific binding.

3.9.2 X-ray structural studies

In order to gain further insight into the binding mode of these fragments, soaking experiments were performed. Fr **21**, **22** and **23** were found soluble in water.
Fr **20** was not soaked due to availability issues at the time of the experiments. Fr **22** was soaked in a range between 10-50 mM in reservoir for 10-24 h but upon inspection of structures, no electron density was detected to fit the ligand.

3.9.2.1 Fragment 21

Fr **21** was soaked at 10-20 mM in buffer crystallization for 10-24h. Higher concentration led to immediate dissolution of the crystals, while longer soaking time (above 24h) were tolerated by the crystals but destroyed the diffraction pattern. Inspection of the structures following data processing led to the identification of extra electron density only for BAZ2B PHD crystals.
This was the first structure of BAZ2B PHD solved with a ligand bound and the first structure of a PHD with a fragment bound in the histone pocket.
Structure was obtained upon soaking of the crystal for 1 day at 20 mM concentration of Fr **21**. Crystals soaked for less than 24 h did not show any fragment bound. Crystallographic data are reported in Table 3.11.

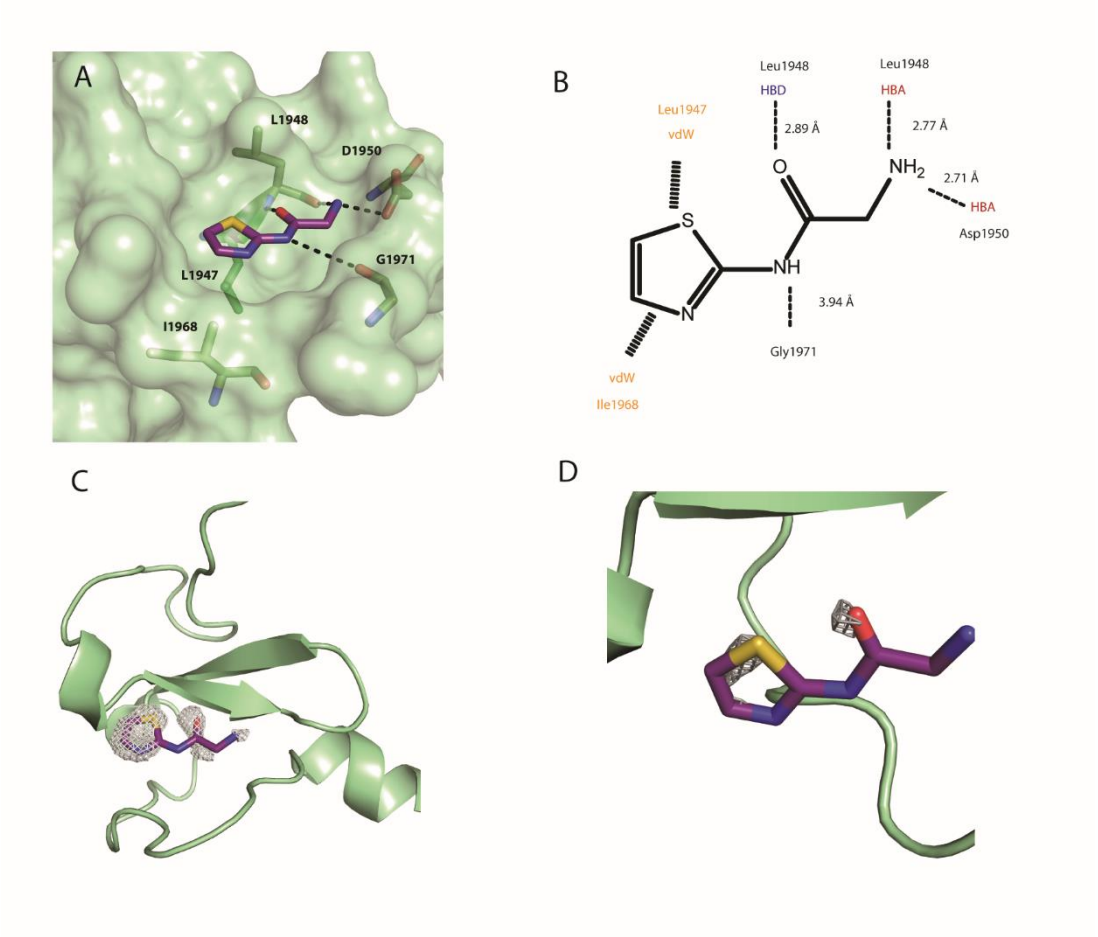


Figure 3.25. Crystal structure of BAZ2B PHD in complex with Fr21.

(A) Crystal structure of BAZ2B PHD (pale green surface) in complex with Fr **21** (purple sticks). Residues interacting with Fr **21** are drawn in sticks. Oxygen, nitrogen and sulphur are respectively reported in red, blue and yellow. (B) Planar representation of the interactions of the fragment with BAZ2B. In Angstrom the distance of the hydrogen bond. HBA stands for hydrogen bond acceptor and HBD for hydrogen bond donor, relatively to the residue on the protein. In (C) and (D), the Fo-Fc map in grey contoured respectively at 2σ and 3.5σ.

Hotspots of the interaction between BAZ2B and Fr **21** are reported in fig. 3.35 A and B. These are in agreement with the hotspots identified by HSQC titration (Figure 3.24). The amino terminal group of the molecule is involved in H-bonds with the side chain of D1950 and the carbonyl group of L1948, while the carbonyl of the amide group of Fr **21** is making hydrogen bond with the NH of L1948. The above interactions are conserved with those observed in the BAZ2A-Fr **19** structure (further discussed below). The NH of the amidic group of the molecule may be involved in electrostatic interactions with the carbonyl group of the G1971. The thiazole group was modelled with the sulphur atom

pointing towards the beta strand of the PHD since Fo-Fc map showed density up to 3.5 σ level (figure 3.25 D). The aromatic group is located just above the hydrophobic pocket, making interactions with the side chains of L1947 and I1968.

Superposition of BAZ2B-Fr 21 with BAZ2A-Fr 19.

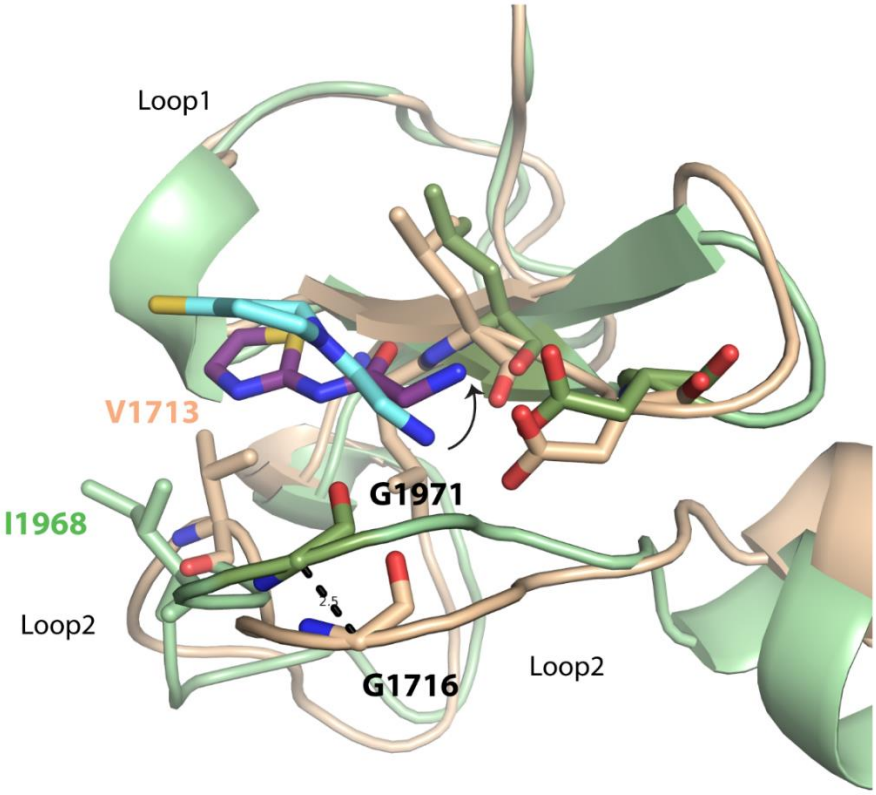


Figure 3.26. Superposition of the crystal structure of BAZ2B PHD in complex with Fr 21 and BAZ2A PHD in complex with Fr 19.

Superposition of the crystal structure of BAZ2B (cartoon, in palegreen) in complex with Fr **21** (violet sticks) and BAZ2A (cartoon, in wheat) in complex with Fr **19** (cyan, sticks). Oxygen, nitrogen and sulphur are respectively reported in red, blue and yellow. The arrow indicates the rotation of the 2-amino acetamide plane.

Superposition of the complex BAZ2B-Fr **21** and BAZ2A-Fr **19** (Figure 3.26) highlighted that the binding mode is mainly conserved between the two proteins, indeed, they are involved in hydrogen bonds with the same set of residues. The introduction of a secondary amide with Fr **21** stabilizes the interaction of the aromatic group with the hydrophobic pocket. A major difference is the 2-amino acetamide plane which is

oriented in a different manner in the two fragments, making an angle of 69.7° between the two. In Fr **21** all the torsion angles lie on the same plane of the aromatic ring, consistent with extensive delocalization of the π system. This particular structural orientation favours the electrostatic interaction of the NH of the amide of Fr **21** with the backbone carbonyl of G1971, while the corresponding conserved residues in BAZ2A (G1716) point towards the amino terminal group of Fr **19** (Figure 3.26).

The orientation of the ligand could be influenced from the different position of the loop2 in the two proteins. Indeed, the measured distance between the Cα of the two conserved residues G1971 (BAZ2B) and G1716 (BAZ2A) is 2.5 Å. Also the hydrophobic pocket has a more open conformation in BAZ2B due to the residues I1968 pointing outside of the protein, instead the correspondent V1713 in BAZ2B points toward the protein itself (Figure 3.26).

Another scenario that may influence the conformation adopted from Fr **21** in BAZ2B is the presence of a crystal contact. Indeed, in the pocket we observe the presence of the side chain of D1941, from an adjacent ASU that, through electrostatic interaction with the carbon of the amide of Fr **21**, could help the stabilization of the small molecule in that particular position (figure 3.27).

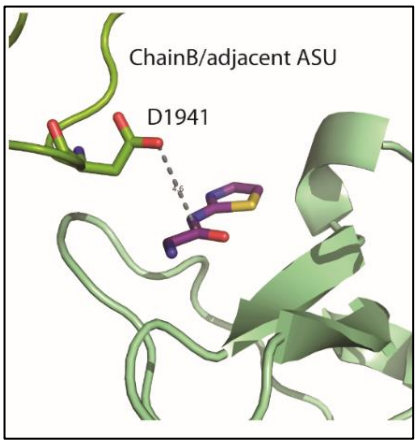


Figure 3.27. Crystal contact with adjacent ASU.

Crystal contact between the carbon of the carbonyl group in Fr **21** and the oxygen of the side chain of D1941 of an adjacent ASU.

Superposition of BAZ2A-H3 3-mer with BAZ2B-Fr 21.

Analysis of the superposition of the complex of BAZ2A with H3 ART and Fr **21**, can

underline interesting features that may be of importance towards the elucidation of the binding mode and to assist future optimization of the leading fragment. The overall area occupied by the fragment is comparable with the area occupied by the backbone of the H3 tripeptide. Interestingly, the fragment in BAZ2B adopt a different orientation from the tripeptide ART. The NH_2 group of Fr **21** displaces the water molecule W1, located between the side chain of D1950 and the carbonyl of L1948 (Figure 3.28). This water is necessary for the binding with the peptide, as described in Chapter 2 (Figure 2.2). The carbonyl group of Fr **21** is involved in H-bond with the backbone of L1948, in a similar fashion to the carbonyl of R2. The thiazole group is in proximity of the hydrophobic region occupied by the methyl group of the side chain of T3.

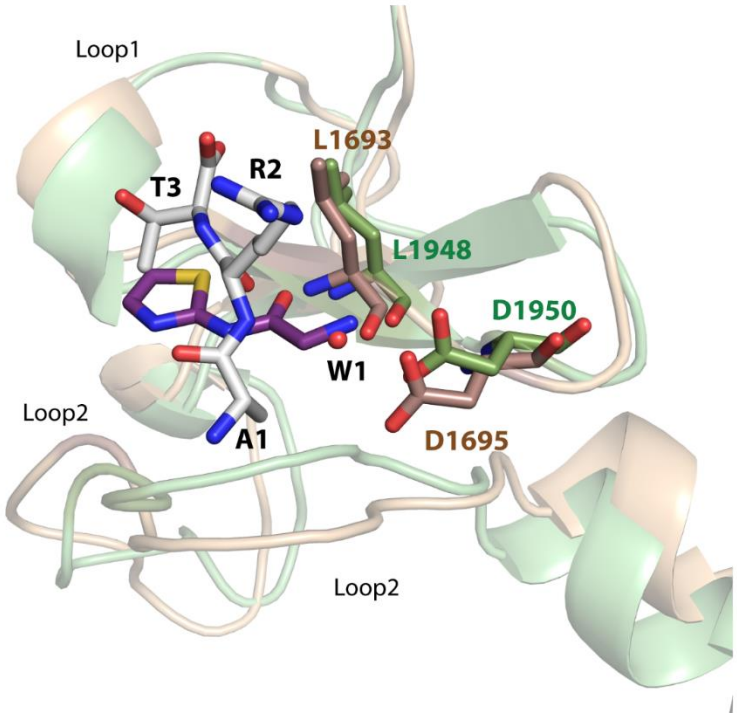


Figure 3.28. Superposition of the crystal structure of BAZ2B PHD in complex with Fr 21 and BAZ2A PHD in complex with ART.

Superposition of the crystal structure of BAZ2B PHD (cartoon, in pale green) in complex with Fr **21** (violet sticks) and BAZ2A PHD (wheat cartoon) in complex with ART tripeptide (grey sticks). Oxygen, nitrogen and sulphur are respectively reported in red, blue and yellow. Water W1 is reported as red dot.

3.9.2.2 Fragment 23: structural studies of binding mode

Fr **23** was soluble in water up to 1 M concentration. Soaking fragments up to 50 mM and up to 24 h was well tolerated from crystals of BAZ2A and BAZ2B which both assumed a

brown colour typical of the fragment solution (figure 3.29).

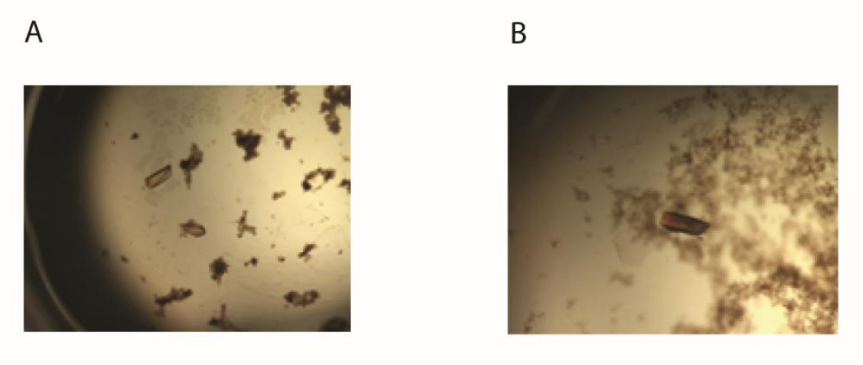


Figure 3.29. Appearance of crystals of BAZ2A/B after 24h soaking with Fr 23.

Crystals of BAZ2B PHD in (A) and of BAZ2A PHD in (B).

Structures of PHD of BAZ2A/B were solved as described in chapter 6.5 and refinement statistics are reported in the Table 3.11.

Inspection of the electron density identified the ligand presence in the histone pocket of both PHDs. Extra electron density in the histone pocket of two chains, B and C was observed for BAZ2A protein, and in one chain for BAZ2B protein. In figure 3.30 and 3.31, the fragment bound structure of BAZ2A and BAZ2B with Fr **23** solved at 2.7 Å resolution.

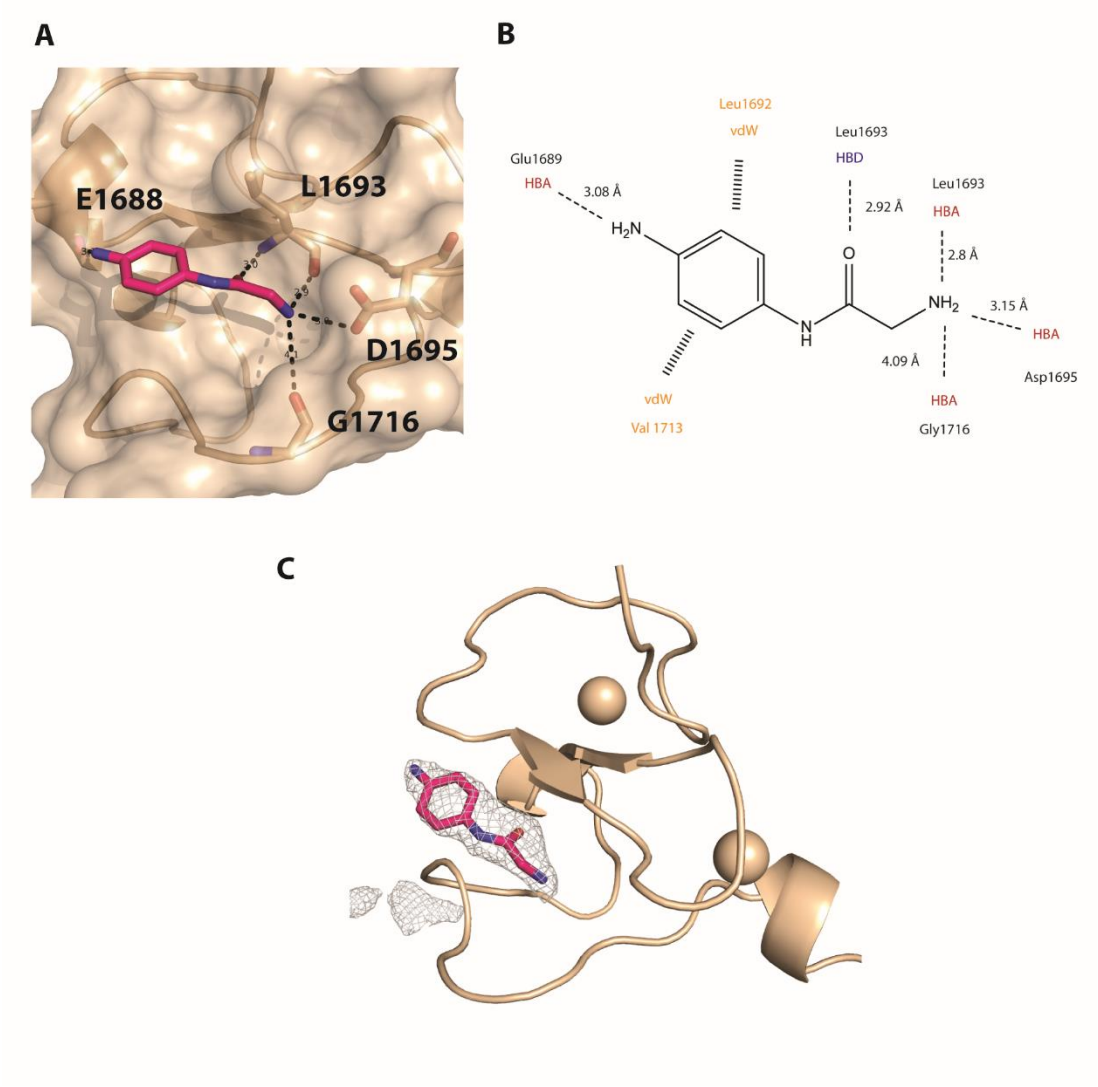


Figure 3.30. Crystal structure of BAZ2A PHD in complex with Fr 23.

(A) Crystal structure of BAZ2A PHD (pale green surface) in complex with Fr **23** (pink sticks). Residues interacting with the fragment are drawn in sticks. (B) Planar representation of the interactions of the fragment with BAZ2A. Distance of the hydrogen bonds are in Å . HBA stands for hydrogen bond acceptor and HBD for hydrogen bond donor, relatively to the residue on the protein. (C) F_o-F_c map in grey countered at 2σ.

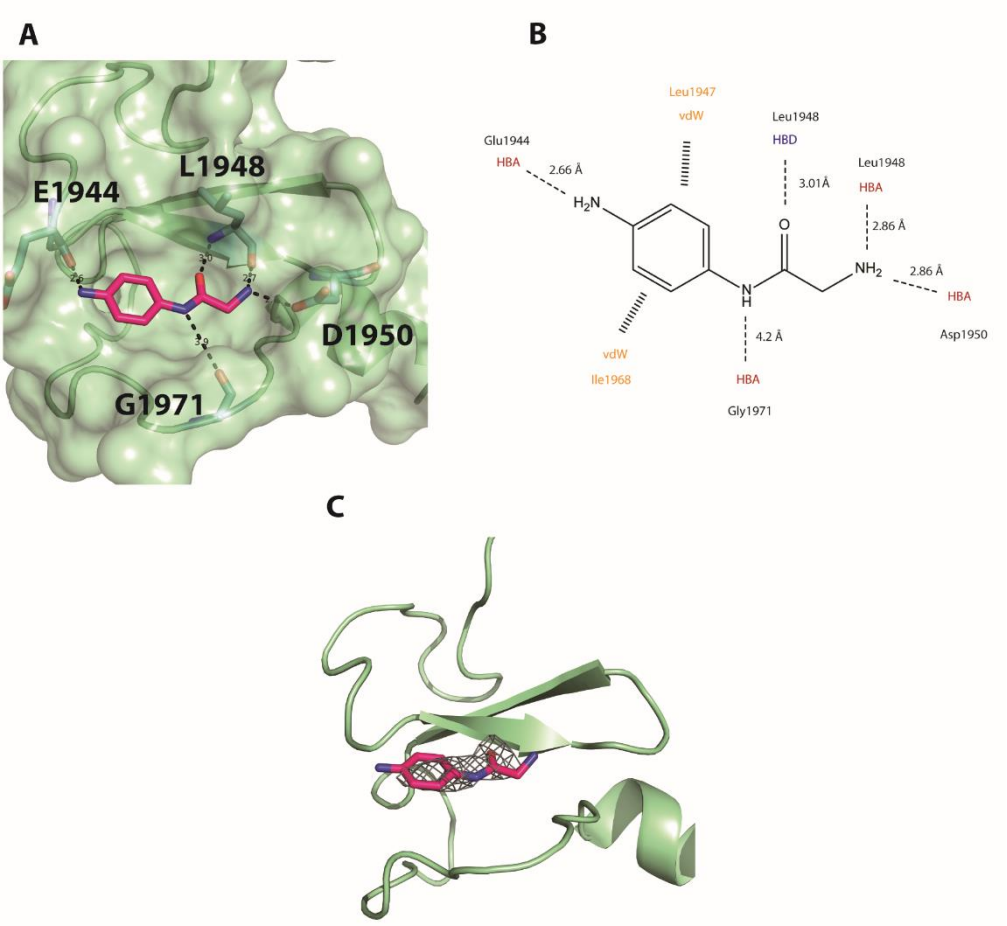


Figure 3.31. BAZ2B PHD in complex with Fr 23.

A) Crystal structure of BAZ2B PHD (pale green surface) in complex with Fr **23** (pink sticks). Residues interacting with the fragment are drawn in sticks. (B) Planar representation of the interactions of the fragment with BAZ2B. In Å the distance of the hydrogen bonds. HBA stands for hydrogen bond acceptor and HBD for hydrogen bond donor, relatively to the residue on the protein. (C) F_o-F_c map in grey countered at 2σ.

Figures 3.30 and 3.31 highlights the hotspots of the interaction, which show to be consistent between the two proteins and in agreement with the heat map obtained from HSQC titration of both proteins (figure 3.24). The amino group on the aromatic ring points towards hydrogen bond formation with the backbone of E1689 in BAZ2A and of E1944 in BAZ2B, while we expected to catch interactions with the side chain. The fact that we are not able to observe this interaction in the crystals does not mean that is not happening in solution, indeed, the particular crystal form adopted could be a limitation that does not allow the movement of the side chain towards the aniline group.

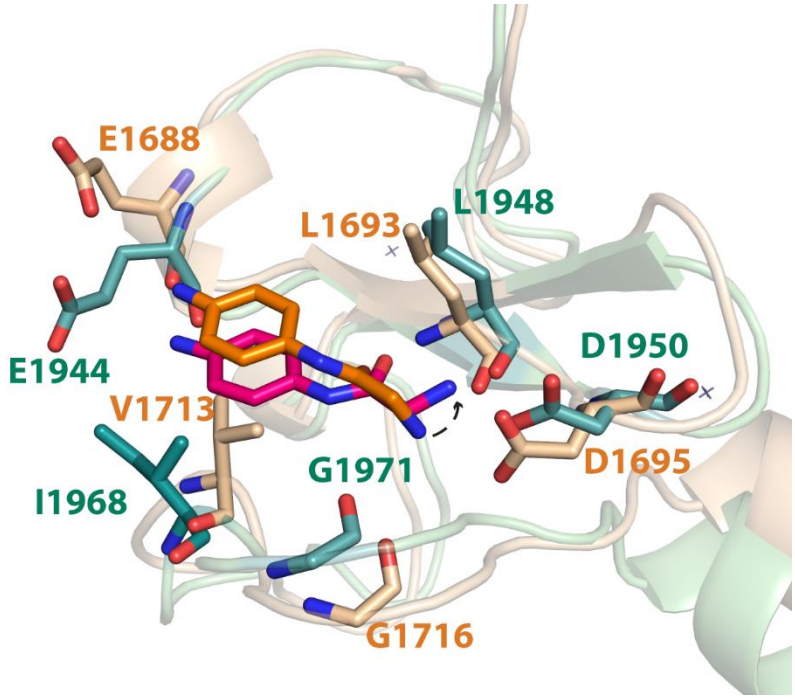


Figure 3.32. Superposition of the crystal structure of Fr 23 bound to PHD BAZ2A/B.

Superposition of the crystal structure of BAZ2B (cartoon, in palegreen) in complex with Fr **23** (pink sticks) and BAZ2A (cartoon, in palegreen) in complex with Fr **23** (orange, sticks). The arrow indicates the different orientation of the terminal amino group.

Key differences are observed in the binding mode of the two proteins. Firstly, the 2-amino acetamide planes assume a different orientation. In Fr **23** bound to BAZ2B the torsion angle of the acetamide plane is planar to the aromatic ring, while in Fr **23** bound to BAZ2A the torsion angle is clockwise rotated of 46° compared to the plane of the ring. This rotation is needed to retain the hydrogen bond with the protein despite the aromatic ring being pushed further by the presence of V1713 in a closer conformation in BAZ2A than BAZ2B (Figure 3. 32). Indeed, the orientation of the 2-amino acetamide plane in Fr **23** bound to BAZ2A is retained also in Fr **19** in complex with BAZ2A (figure 3.33).

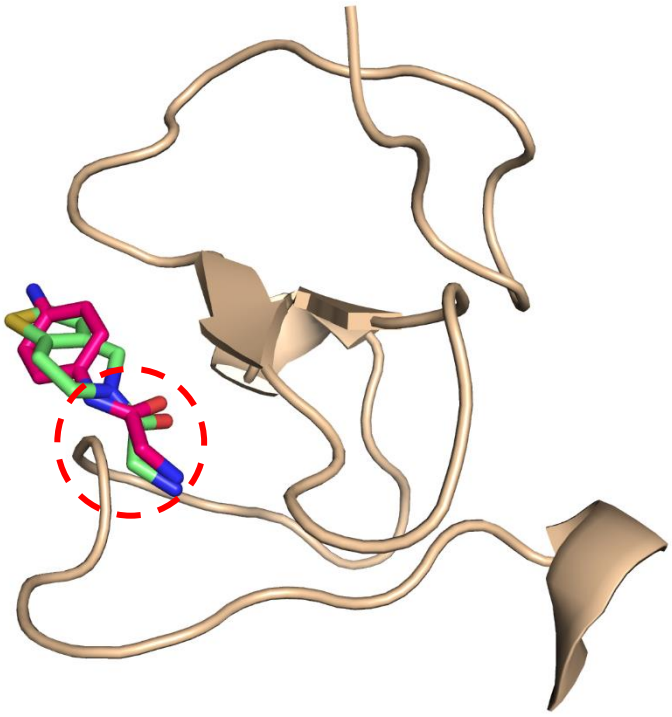


Figure 3.33. Superposition of the crystal structure of Fr 23 and Fr 19 bound to PHD BAZ2A.

Superposition of the crystal structure of BAZ2A (wheat cartoon) in complex with Fr **23** (pink sticks) and Fr **19** (cyan sticks) showing that the 2-amino acetamide retains the orientation in both fragments (highlighted by red dashed circle).

This difference in the torsion angle accounts for some differences in the binding mode between the two proteins. The H-bond distances in BAZ2B are shorter and we observe an electrostatic interactions between G1971 and the amide of the ligand (figure 3.32 and 3.31), while in BAZ2A, the correspondent glycine residues is involved in electrostatic interactions with the N-terminal amino group of the ligand (figures 3.30 and 3.32).

Comparison between binding mode of Fr 23 and ART peptide

The superposition of the complex of BAZ2A with Fr **23** and ART peptide (Figure 3.34) revealed molecular details previously highlighted in the superposition with Fr **21**, in figure 3.28.

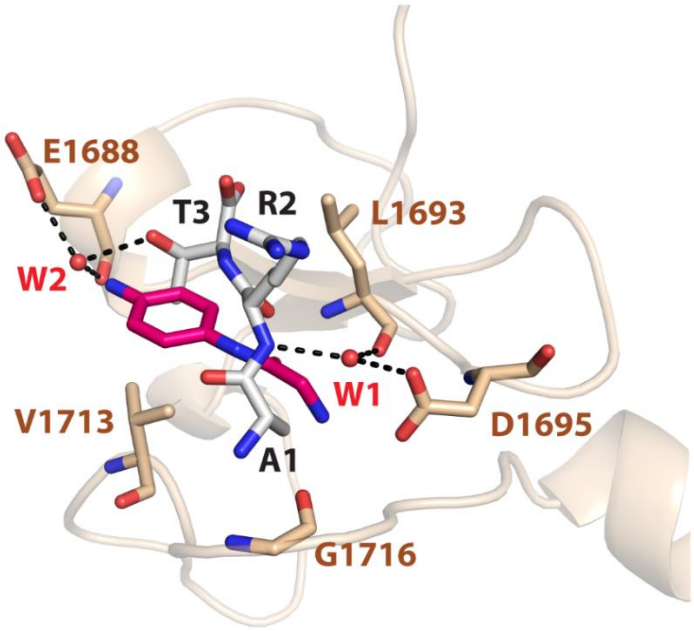


Figure 3.34. Superposition of complex of BAZ2A with Fr 23 and ART tripeptide.

Superposition of the crystal structure of BAZ2A PHD (wheat cartoon) in complex with Fr **23** (pink, sticks) and ART tripeptide (grey sticks). Waters are reported as red dots and bond distance with black dashes.

3.9.2.3 Summary of the crystallographic data

Table 3.10 summarizes the number of structures obtained for each fragment and 3-mer peptide and the number of ligands found in each asymmetric unit. Table 3.11 A and 3.11 B display the crystallographic data of the structures solved.

Protein	Protomers in ASU	ART 3-mer peptide bound	Fr19 bound	Fr21 bound	Fr23 bound
BAZ2A PHD	4	3/4	2/4	X	2/4
BAZ2B PHD	2	X	X	1	1

Table 3.10. Summary of the peptide and fragment bound structures and number of ligand for each ASU. 'X' denote no fragment or peptide bound.

Data collection	BAZ2A-ART	BAZ2A-Fr 19	BAZ2A-Fr 23
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$
Cell dimensions			
a, b, c (Å)	73.4, 73.4, 99.7	72.1, 72.1, 99.2	72.8, 72.8, 99.6
α, β, γ (°)	90, 90 , 90	90, 90 , 90	90, 90 , 90
Resolution (Å)	46.03 (2.0)	45.34 (1.9)	45.2 (2.7)
Unique observations	19049	21069	7841
Completeness	99.9	99.3	100
Redundancy	8.6 (9.0)	3.8 (4.0)	12 (12.3)
CC1/2	0.99 (0.78)	0.99 (0.83)	0.99 (0.95)
I/σI	15.5 (2.1)	14.3 (2.6)	10.7(3.7)
Refinement			
R_{work}/R_{free}	19.6/23.5	20.3/23.7	17.4/22.4
B factor ($Å^2$)	35.3	36.7	62.7
Rmsd bond (Å)	0.011	0.01	0.005
Rmsd angle (°)	1.6	1.7	0.75
Ramachandran statistics			
Allowed (%)	0.5	0.5	2.5
Favoured (%)	99.5	99.5	97.5
Outliers (%)	0	0	0

Table 3.11.A. Crystallographic data of the BAZ2A PHD structures.

In parentheses data relative to highest resolution shell.

Data collection	BAZ2B-Fr 21	BAZ2B-Fr 23
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions		
a, b, c (Å)	38.3, 45.4, 65.0	37.8, 45.5, 64.8
α, β, γ (°)	90, 90 , 90	90, 90 , 90
Resolution	65.04 (1.95)	64.83 (2.7)
Unique observations	8453	3349
Completeness	97.7 (87.1)	100.0 (100.0)
redundancy	5.8 (4.9)	5.6 (5.9)
CC1/2	0.99 (0.91)	0.96 (0.56)
I/σI	18.1 (1.3)	5.3 (1.9)
Refinement		
R_{work}/R_{free}	18.5/23.8	22.6/29.9
B factor ($Å^2$)	34.1	67.7
Rmsd bond (Å)	0.02	0.014
Rmsd angle (°)	1.88	1.73
Ramachandran statistics		
Allowed (%)	2.7	5.5
Favoured (%)	97.3	94.5
Outliers (%)	0	0

Table 3.11.B. Crystallographic data of the BAZ2B PHD structures.

In parentheses data relative to highest resolution shell.

3.10 DISCUSSIONS

In this chapter, I presented extensive biophysical and structural validation of six out of 19 fragments identified from an *in silico* screening. The K_D and potential region of binding were determined for some of the fragments using CSPs. One of the validated fragments, Fr **19**, led to a co-crystal structure with BAZ2A PHD, although the electron density did not cover the entire small molecule. This first fragment-bound structure identified a key recognition site at the histone binding pocket, consistent with the chemical shift perturbation data observed by HSQC NMR.

These co-crystal structures guided further optimization of the fragment in order to improve the binding mode, by purposefully targeting the hydrophobic region of the histone pocket and the acidic wall. Four fragments were designed as outcome of this optimization and their binding mode was further characterized by HSQC and X-ray crystallography. Two of these fragments were successfully soaked in BAZ2A and BAZ2B PHD. I solved the structure of Fr **21** with BAZ2B and Fr **23** with both BAZ2A and BAZ2B PHDs. All fragments bound to the same conserved hotspots, in agreement with the HSQC data. Analysis of these structures led to the identification of similarities and differences in molecular recognition between the two proteins, as well as similarities and differences in binding mode between the fragments and the histone peptide.

In summary, this work furnishes a blueprint for the development of future improved ligands of this family of PHDs: BAZ2A and BAZ2B. All fragments recapitulate a conserved hotspot at the histone binding site, which however is not explored by the histone tail. It therefore may be possible to combine information from the two binding modes e.g. by merging features of the two types of molecules, to enhance the yet moderate binding affinity of these compounds. Further exploration of the peptide binding features could contemplate a rational design to take advantage of the important histone tail helicity revealed in the previous Chapter, for example by linking our identified fragments with synthetic scaffolds able to mimic or recapitulate those alpha helical features. Such approaches could aid the development of novel chemical probes disrupting the interaction between BAZ2A/B PHD finger domains and histone H3, which could

ultimately provide new chemical tools to interrogate the still elusive molecular structure, activity and biological function of these proteins.

CHAPTER 4.
STRUCTURAL STUDIES AND DE-ORPHANIZATION OF PHD-BROMODOMAIN TANDEM MODULES

4.1 MOTIVATION OF THE WORK

PHD finger- bromodomain tandem modules characterise several chromatin-associated proteins and are generally known to interact with histone proteins. Mutations associated with their loci are often linked to tumorigenesis and genetic disorders [39, 55]. Since their involvement in the development of human disease, these reader domains are regarded as potential targets for drug development [36, 43, 70].

This chapter focuses on the identification of the histone binding partner of the PHD finger – bromodomain of BAZ1B and TRIM66 with the aim to elucidate, at molecular level, the basis of this interaction. Although it is known from literature that both proteins are involved in diseases [108, 115, 116, 140], the structure-function properties of their PHD-bromodomain modules remain elusive, and no information have been gained yet about their substrate binding partner. Indeed, if it is possible to associate binding of an acetyllysine residue on a histone tail to the bromodomain, the binding partner of the PHD zinc finger is less predictable since these domains can recognise a much broader range of substrates and PTMs [46]. Moreover, it was noticed that these two tandem modules are both present at the C-terminus of the protein but the linker inter-connecting the two reader domains is substantially different in length, suggesting a potential different histone recognition mode for the two tandems (either in *cis* or *trans*).

The main goal of the work here presented is to provide structural information and elucidate the molecular recognition features of these domains. In doing so, we aim also to develop a platform of biophysical and structural assays to assist the future design of small molecules able to target these reader domains. The future ambition will be to furnish a chemical scaffold that could be further developed either into an inhibitor or into a probe useful to investigate the biological function of the proteins themselves.

4.2 BAZ1B ALSO KNOWN AS WSTF

The strategy adopted to study the PHD-BrD tandem module of BAZ1B was to dissect the tandem in the two individual reader domains, PHD zinc finger and bromodomain, which were investigated first as independent units. Here, there are described the efforts toward crystallization of BAZ1B PHD, which however proved unsuccessful, and several attempts to identify the histone binding partner of this PHD finger.

4.2.1 Insights into deorphanization of PHD of BAZ1B

The BAZ1B PHD construct (figure 4.1) was designed based on the constructs of PHDs of the same BAZ family, previously studied in the lab (BAZ2A/B). It was cloned into the pCri11b plasmid which allows expression of BAZ1B PHD fused, at its N-terminus, with a 6His-SUMO tag. The expression of the BAZ1B PHD was performed in BL21(DE3) *E.coli* strain and the purification followed four steps, described in details in the Experimental section.

The first step was a Ni^{2+} affinity chromatography, followed by overnight cleavage of the His_6-SUMO tag and a further step of affinity chromatography for the tag removal. At this point, pooled fractions from the previous step were dialyzed to reduce the salt concentration in the buffer, and the protein was further purified using anion exchange chromatography. The last polishing step was size exclusion chromatography (SEC) in buffer supplemented with 20 μM $ZnCl_2$ to assure native conditions to the sample, given the protein chelates two zinc ions (figure 4.1). The final yield obtained was 10 mg per litre of culture and the right MW was confirmed by ESI-MS (figure A.1 Appendix).

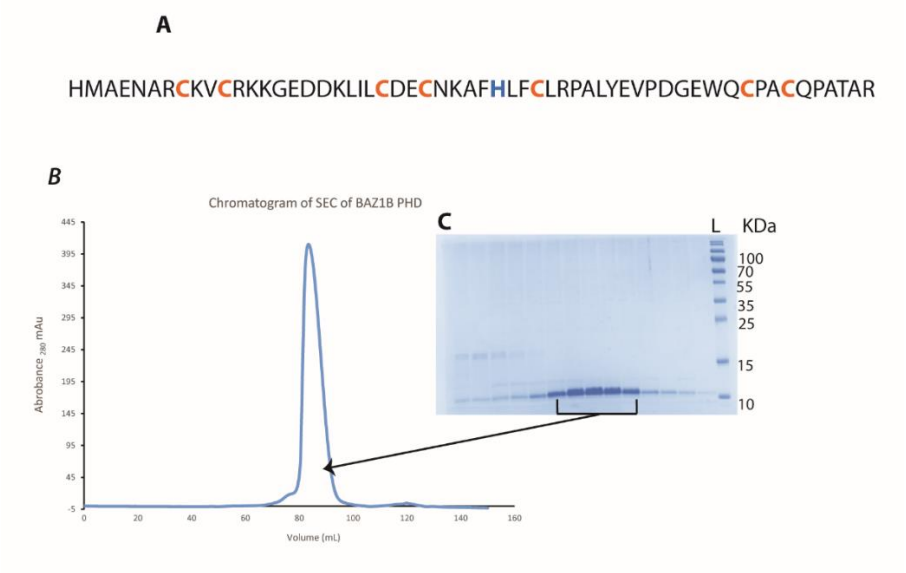


Figure 4.1 Purification of PHD zinc finger domain of BAZ1B.

(A) Amino acid sequence of PHD of BAZ1B. Highlighted in orange the Cys and in blue the His residues belonging to the conserved motif Cys4-His-Cys3 that coordinates the two zinc ions. (B) Gel filtration elution profile of PHD of BAZ1B. In blue, the protein absorbance. (C) SDS-page: from left to right the fractions under the peaks with the protein ladder on the right.

1D proton NMR spectrum provided a rapid evaluation of the protein folding.

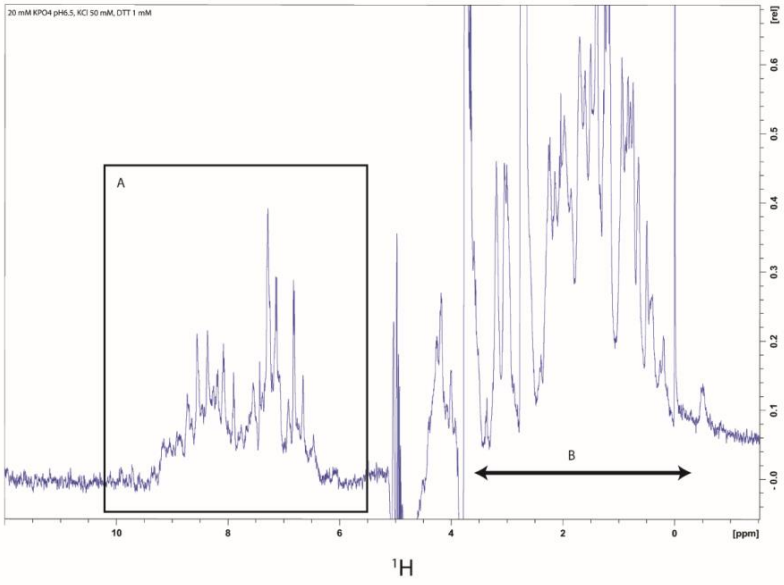


Figure 4.2 1D NMR spectrum of PHD of BAZ1B.

In A and B the details of the two most interesting regions of the spectrum.

In a spectrum of folded protein, the signal is scattered on a large range of chemical shifts and the peaks appear sharp and narrow because the protons are constrained in a specific conformation and the environment of each proton is influenced by the tertiary structure of the protein. In an unfolded spectrum we would see a set of peaks not so widely dispersed and with a broad shape; the reason is that in an unfolded protein the protons can experiment different environments and the signal is just an average of them.

The most meaningful part of the spectrum is the region indicated with A in fig. 4.2. It represents the amide protons region, between 7-10 ppm, which also contains aromatic protons. In B there is the aliphatic region. In Fig. 4.2, the signals are well spread on all the spectrum and the peaks are sharp, so it is possible to infer that the protein is folded.

4.2.1.1 Crystallographic efforts to solve the apo form structure of PHD of BAZ1B

The NMR structure of the apo form of PHD of BAZ1B was already available when I started my PhD. Despite this, it was deemed important to obtain a crystal structure of the apo protein since it could have been used, in soaking experiments, to study, at molecular level, interactions with its potential binding partner.

Several crystallization screening were performed using sparse matrix and grid screens at 4°C and 20°C with different concentration of proteins, in a range between 3 and 14 mg/mL. Crystallization drops were visually inspected and most of them resulted in precipitate or clear drops. It was noticed that in the PACT screen (Qiagen), upon examination, all drops were clear, suggesting the protein is highly soluble in these conditions and saturation was not achieved. Two conditions were chosen to perform some buffer exchange experiments. Indeed, there are examples, in literature, showing how conditions that increase the solubility of the protein can be beneficial to induce crystallization [160]. Two conditions were chosen from the screening: C7 (200 mM NaCl and 100 mM Tris pH 8) and D7 (200 mM NaCl and 100 mM HEPES at pH 7) where drops appeared to be extremely clear. The PHD of BAZ1B was buffer-exchanged in these two empirically identified enhanced-solubility buffers, concentrated to 12 mg/mL and rescreened against the PACT conditions. The outcome was the formation, after few

days, of microcrystals in many drops of the plate. In order to decrease the nucleation points, the protein/crystallization buffer ratio was decreased, but also in this case no hits were obtained.

4.2.1.2 *In situ* proteolysis and optimization of the construct

A factor that can influence the rate of success of the crystallization process is the choice of construct. Indeed, it is important to avoid the presence of flanking regions which might constitute an entropic and steric barrier toward the formation of crystal packing. Indeed, analysis of the crystal structures of the homologous PHD of BAZ2A and BAZ2B, previously solved in our lab [46], showed that the flexible N-terminal tails of these PHDs become ordered in the crystal packing. These N-terminal tails are important for the formation of homodimeric anti-parallel β-sheet contacts in the crystal packing. However, it may be that for PHD of BAZ1B, this hypothetic flexible extension are an obstacle for the crystallization process.

An *in situ* preotelysis [161] approach for crystallization was attempted to potentially overcome the problem of having flexible tails in the protein. Minimal traces of protease were added to the crystallization drops in the screen (protease/protein ratio~1/5000). The choice of the protease was guided by the cleavage site and by enzyme availability. Papaine was the elected one, with the cleavage site that includes a sequence like XX-hydrophobic-X -R/L- X(excluded V), where X is a generic amino acid. Eventually, also this method did not lead to any crystallization hit.

Next, it was decided to design a new construct removing the first four N-terminal residues (His-Met -Ala-Glu) by quickChange PCR. The protein was expressed and purified following the same protocol used for the previous construct, and concentrated to 7 mg/ml. Crystal screen PACT, PegIon, Morpheus and JCSG+ were performed at 20°C. Inspection of the plates identified only formation of microcrystals and precipitate, apart from few clear drops. No improvement in the generation of crystal hits was observed for this shorter construct.

4.2.1.3 Matrix microseeding screening experiments (MMS)

The crystallization process for the PHD zinc finger of the protein BAZ1B has been very challenging and despite the several conditions screened, the crystallization drops inspected so far resulted in precipitate, microcrystals or clear drops. In light of these results, it was decided to adopt another approach: matrix microseeding screening experiments (MMS). It consisted in the usage of crystal seeds, that are small pieces of larger crystals from which it is possible to grow a new crystal, added to the ordinary crystallization screens [162, 163] and not only in the optimization step. To understand how microseeding could help in the crystallization process, it is helpful to refer to the diagram of the crystallization process in figure 4.3, adapted from Stewart *et al*. [164]. Crystallization is a process influenced by several factors: precipitants, protein concentrations, temperature, etc. In a simplistic case, we can interpret this phenomenon as dominated by two main factors: protein concentration and precipitant concentration (figure 4.3).

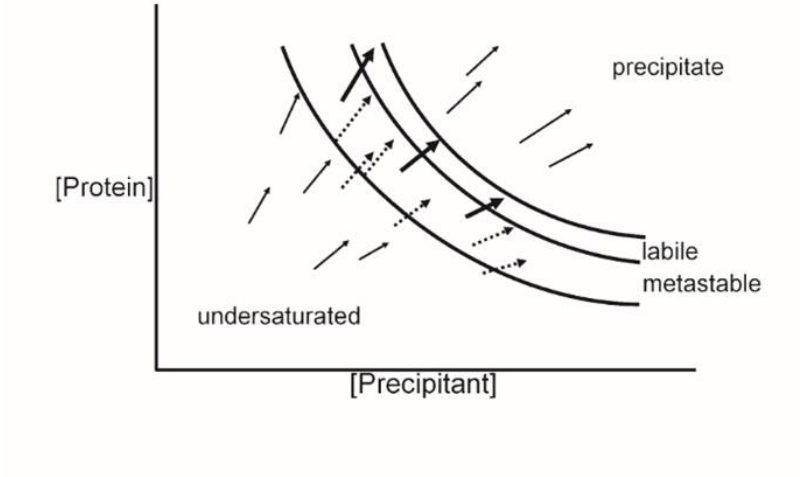


Figure 4.3. Phase diagram of the crystallization process. Figure adapted from Stewart *et al*.[164]
Schematic phase diagram reporting four main areas: undersaturated, metastable, labile and precipitation zone. The arrows are vectors representative of the water evaporation process in the vapor diffusion setup. The initial and ending points of the arrow are representative of the initial and final conditions of the process. Thick arrows delineate the process of crystallization of those conditions that overcome the metastabile zone ending up in the labile zone. Dashed arrows represent conditions that end up in the metastabile zone, while thin arrows depict conditions for which no crystals are observed because too high (precipitation) or too low (undersaturation) concentration of the components.

At high concentration of both components, precipitation is occurring and protein is aggregating in an amorphous way. Lowering one of the two components, we fall in either nucleation (labile) zone or the metastable zone. The nucleation zone is where *nuclei* form and crystals will grow. The ideal situation is the formation of a few nuclei so the majority of the protein present can increase the size of the nucleus giving proper single crystal. In the metastable zone, crystals are only allowed to grow but there are no conditions for the nucleation to initiate. If a seed is added in the metastable zone, it will grow larger giving a crystal. The line between the metastable and labile zone is representative of the solubility of the protein. Instead, in the undersaturated zone, conditions stay clear and no crystals can grow. Difficult in crystallizing often comes when the nucleation zone (labile zone) is very small and there are really few conditions that can lead to the formation and growth of these nuclei. In situation like this, the addition of seeding in the metastable zone can assist the crystal formation.
In literature, there are several cases of proteins where the MMS method performed even with homologous proteins, has notably increased the number and types of hits where little or no nucleation is normally observed [162, 165, 166].

I decided to perform MMS using cross-seeding with seed of the homologous PHDs of BAZ2A and BAZ2B [46, 167]. Crystallization screenings were set up following the protocol from Douglas Instruments (http://www.douglas.co.uk). For the preparation of the microseeds stock the Seed Bead kit provided by Hampton Research was used, following the instructions of the manufacturer. Unfortunately, also random microseeding screening did not provide any interesting hit for the PHD of BAZ1B, but only microcrystals and precipitate.

4.2.1.4 Lysine methylation

A potential barrier to the crystallization of BAZ1B PHD could be the high flexibility of some regions of the protein that might be detrimental to the formation of a well-ordered crystal lattice, as for example the sequence comprising residues 1192-1198 that is enriched in K residues and is shown to be highly flexible in the NMR structural ensemble (Figure 4.4).

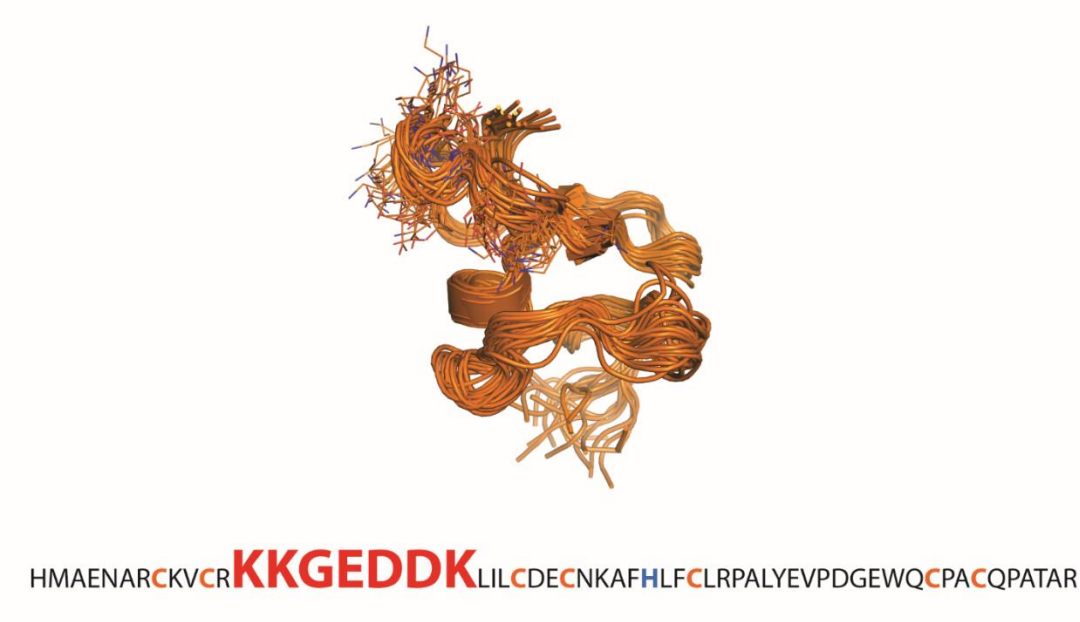


Figure 4.4. NMR ensembles of BAZ1B PHD.

Cartoon representation in orange of the NMR ensembles of the PHD of BAZ1B. In sticks are reported the residues belonging to the area highlighted in red on the protein sequence and enriched of lysine.

Protein surface engineering by chemical modifications is an interesting approach for the reduction of the protein surface entropy due to the mobility of solvent-exposed residues. Indeed, it is a procedure that requires less time and labor compared to the more classical surface entropy reduction by mutagenesis [168, 169]. A method that has proven to be successful on proteins refractory to the crystal formation, is the chemical modification of highly entropic residues, as for example, methylation of solvent exposed lysine [168].

Lysine methylation of the PHD of BAZ1B was performed following the protocol from Walter *et al*. [168]. The MALDI-TOF mass spectrometry confirmed that all the Lysine residues were dimethylated and this was a clear signal that all of them were on exposed surface (figure A.3 in Appendix). The dimethylated BAZ1B PHD was purified and concentrated to 14 mg/mL. Several crystal screening were performed in sparse matrix and grid screens at 4°C, 15°C and 20°C. Unfortunately, no evidence of crystal formation was observed.

All these failures in the crystallization of the apo form of the PHD of BAZ1B motivated the choice to focus the next efforts on the identification of a possible histone binding

partner for this PHD. Indeed, the discovery of a binding partner might have been crucial for the formation of a stable complex, more liable of crystallization.

4.2.1.5 Identification of the histone binding partner of BAZ1B PHD

Sequence alignment of the PHD of BAZ1B with the PHDs of BAZ2A/B suggests that the binding of BAZ1B PHD to the unmodified histone H3 is conserved, since the majority of the residues in the region that binds the histone (circled with the red box in figure 4.5) are conserved. Moreover, it was noticed from the alignment that this domain lacks of the hydrophobic cage typical of the methyllysine PHD binders [45]. In order to validate this hypothesis, an ITC titration was performed but no meaningful binding was detected towards the wild type H3 21-mer peptide.

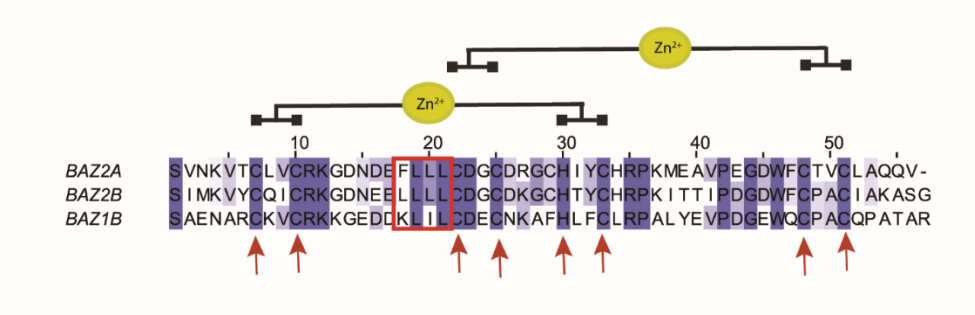


Figure 4.5. Sequence alignment of the BAZ1B PHD with BAZ2A/B PHD.

Sequence alignment was performed using Jalview Software [170]. Red arrows highlight the *Cys4-His-Cys3* conserved motif of the PHD. The red box is highlighting the residues of the β-strand of the PHD that are involved in histone binding. Conserved residues are coloured in blue.

BLI experiments were next performed in order to investigate if PTMs on the histone tail could be necessary for the binding to the PHD of BAZ1B.

BAZ1B PHD protein was screened against a library of 94 different peptides resembling the N-terminal tails of H3 and H4 histones, containing several PTMs. BLI response are reported in figure 4.6.

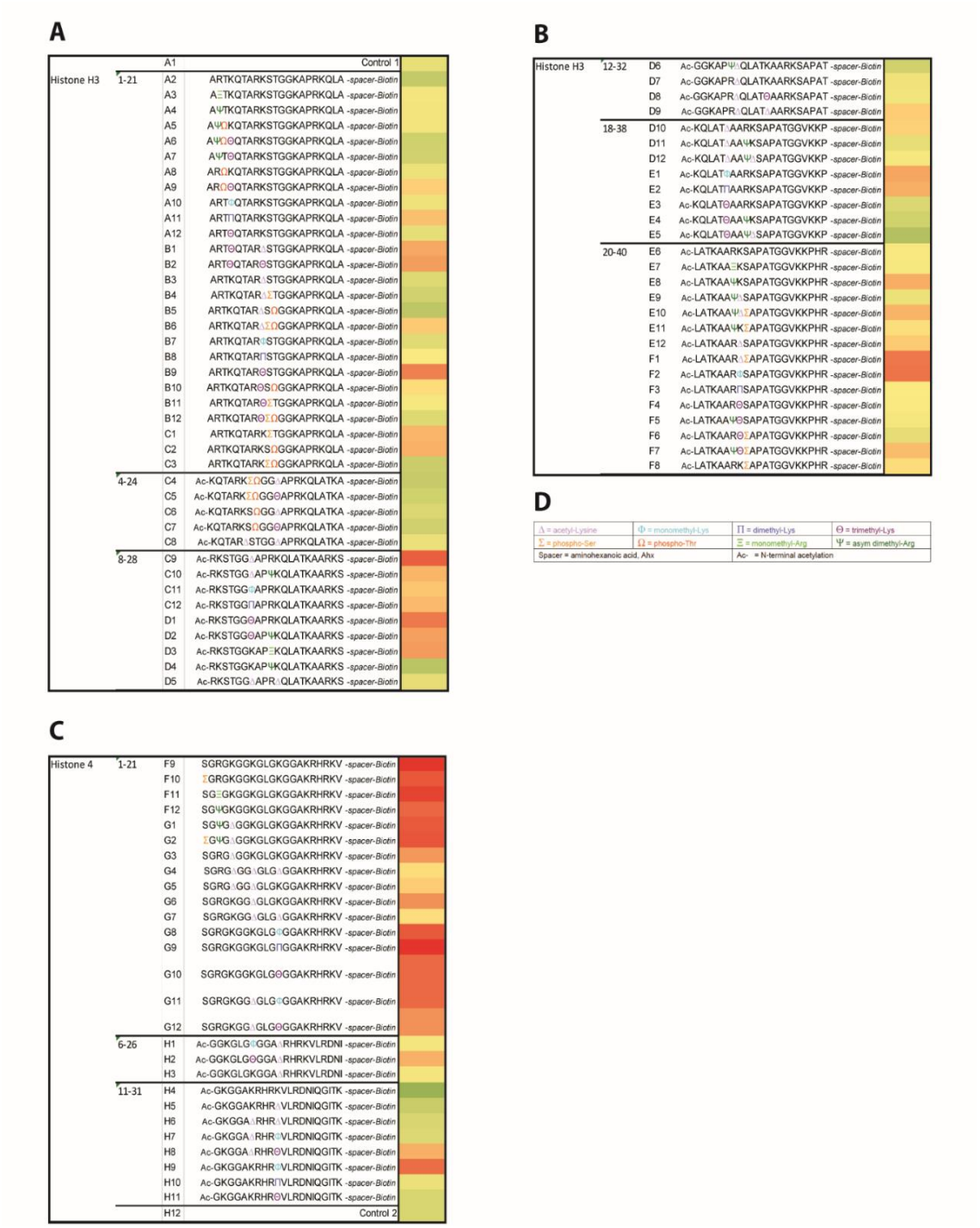


Figure 4.6. BLI responses of BAZ1B PHD towards H3/H4 peptide library.

Biotinylated peptides were immobilized on the surface of the BLI tips while the protein was in the well. Highest response detected are shown in red (0.27 nm for H4 K12me2), lowest responses in green (0.04 nm for the unmodified tail H4 11 -31 aa). Intermediate response are in orange. (A-B**)** BLI responses of PHD of BAZ1B towards H3 tails. (C) BLI responses for the H4 N-terminal tail. (D) Legend for the post translational modifications associated to the H3/H4 tails.

Results showed strong binding preference for the unmodified H4 tail, indeed, in the picture is visible a strong red cluster in correspondence of H4 and its derivatives. Additional modifications, especially acetylation in position H4K5, H4K8 and H4K12, decrease the binding with the exception of the di-methylation in position K12. Furthermore, the response increases again with the methylation on the K20 (H4K20me). Another interesting observation comes out from the sequence pattern around H4K5, K8 and K12; indeed, they share common motifs as: GRG**K_5**GG, K_5GG**K_8**LG and GLG**K_{12}**GG. Acetylation on the Lys in this position was observed to reduce the binding response.

In agreement with the ITC results, no binding was detected for unmodified N-terminal H3 tail. This is in marked contrast with the PHD fingers of BAZ2A and BAZ2B [45]. However, the binding response of BAZ1B for H3 histone slightly increases as a result of tri-methylation in position K9, and acetylation or tri-methylation in position K14.

Simultaneously, PHD of BAZ2B was used as positive control against the same peptide library; conditions set up of the experiments were identical. The results obtained for BAZ2B PHD were equivalent to those previously obtained in the lab [45], as expected. This allowed comparison of the binding responses between the two proteins. Although the intensities of the BLI responses were comparable, remarkable differences were observed in the shape of the binding curves, which can be informative of the type of binding. With BAZ2B PHD, the sensorgrams, after the initial fast rise, reached immediately a stable plateau and the dissociation step went back to the initial baseline (figure 4.7, upper panel). Sensorgrams of BAZ1B PHD showed a biphasic nature with an initial fast rise followed from a slower rate of binding that does not reach saturation (figure 4.7, lower panel). The biphasic character of the curves in BAZ1B was also reflected in the dissociation step, where the initial fast decay was followed by a slower rate without reaching the baseline (figure 4.7 B). This trend is representative of non-ideal heterogenous binding that could be indicative of non-specific binding of BAZ1B PHD to the histone peptides [171].

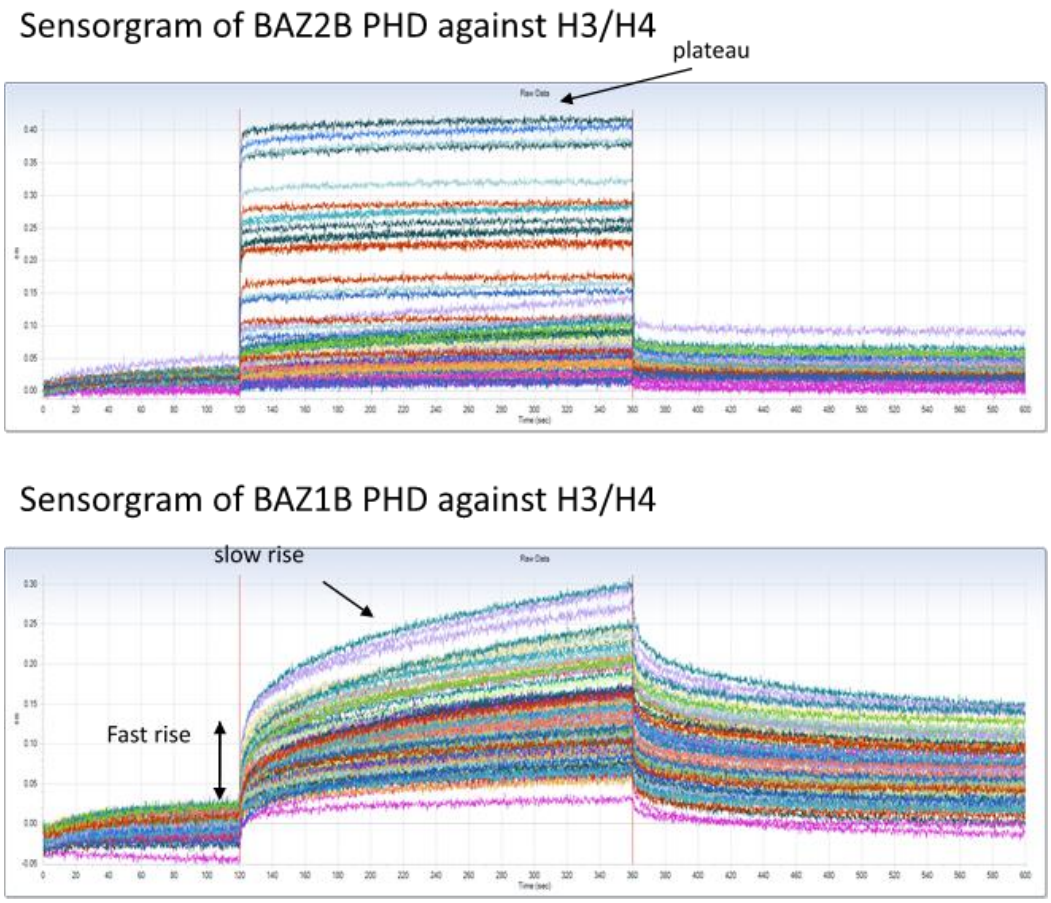


Figure 4.7. Raw BLI sensorgrams obtained for BAZ1B and BAZ2B PHD.

Raw BLI responses for BAZ2B PHD (upper panel) and for BAZ1B PHD (lower panel) towards 94 different biotynilated peptides loaded on the sensor tip. Responses are reported in nm on the y axis over the time on x axis.

The results obtained from this high-throughput assay were subsequently validated by protein-observed NMR, as detailed below.

4.2.1.6 NMR backbone assignment of BAZ1B PHD

To further study the interactions between BAZ1B PHD and its potential binding partners identified in the BLI screening it was decided to use protein-observed NMR experiments. The first step was to assign the HSQC resonances of BAZ1B PHD. Assignment of the PHD of BAZ1B was available in literature [172] but it was here decided to perform our own assignment since we used a longer construct. To this purpose, double labelled $^{15}N/^{13}C$ PHD of BAZ1B was expressed. The level of isotopes enrichment in the expressed protein

was determined by electrospray ionization mass spectrometry (figure A.2 in Appendix) and estimated to be over 95%. Next, sequence-specific backbone assignments were obtained from the identification of intra- and inter-residue connectivities in a series of double and triple resonance spectra: $^{15}N/^{1}H$ HSQC, $^{15}N/^{13}C/^{1}H$ HNCO, HNCACB, HN(CO)CACB. All the backbone amide protons were assigned with the exception of prolines and the N-terminal His (not observable since the terminal NH_3^+ group exchanges with the solvent) (figure 4.8).

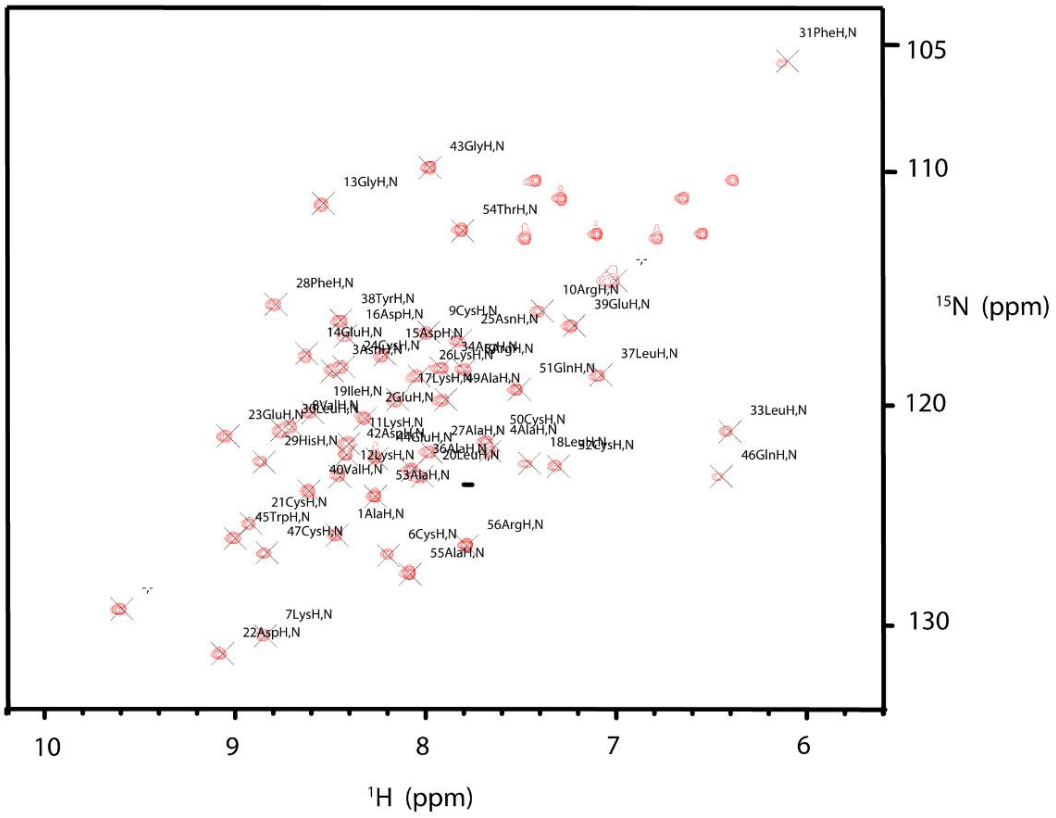


Figure 4.8. HSQC spectrum of BAZ1B PHD with the backbone assignment.

Assigned peaks are labelled in black with the name of the amino acid.

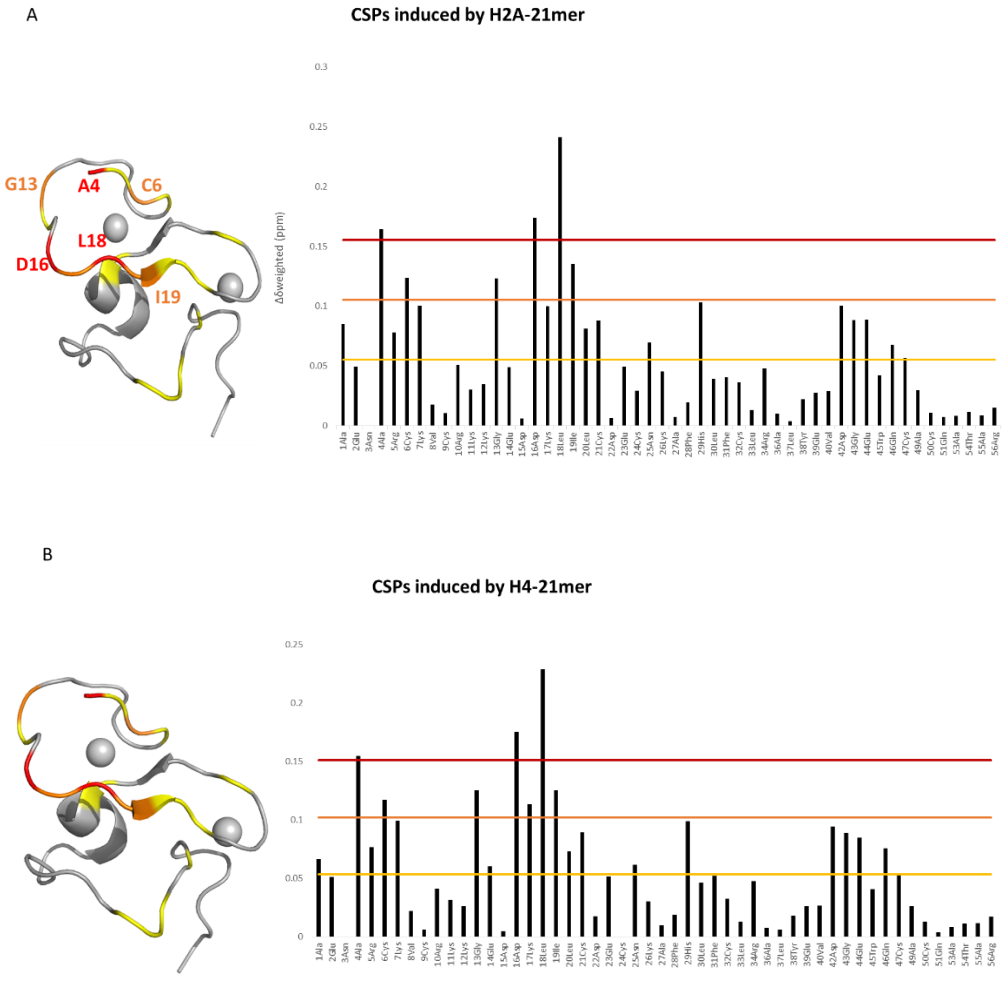
4.2.1.7 CSPs experiments using histones derived peptides

To confirm the validity of the results of the BLI experiments, and eventually describe the region of BAZ1B PHD of binding of the histones, chemical shift perturbation (CSP) experiments were performed using histones derived peptides. A set of four 20-mer histone peptides derived from the unmodified N-terminal tail of H2B, H2A, H3 and H4,

was synthesized. Prior to inserting any specific modification on the histone tail, we tested unmodified histone peptides by CSPs. The reason for this was that in BLI the highest response was detected for the unmodified H4 peptide, consequently, it was expected to be the stronger binder amongst the series.

(^{15}N, ^{1}H)-HSQC spectra were recorded on ^{15}N labelled PHD of BAZ1B with increasing concentration of 20-mer peptide. Δδ were calculated at the highest ratio using the CcpNMR software [154] and plotted for each protein residue on a histogram. CSPs obtained for each peptide were projected onto the BAZ1B structure (PDB: 1F62) by colouring each residue according to the intensity of the shift registered (figure 4.9).

Within the series of the four peptides, highest shifts were reported for H4 and H2A (figure 4.9 A and 4.9 B) that share conserved sequences in the first ten residues (figure 4.10).



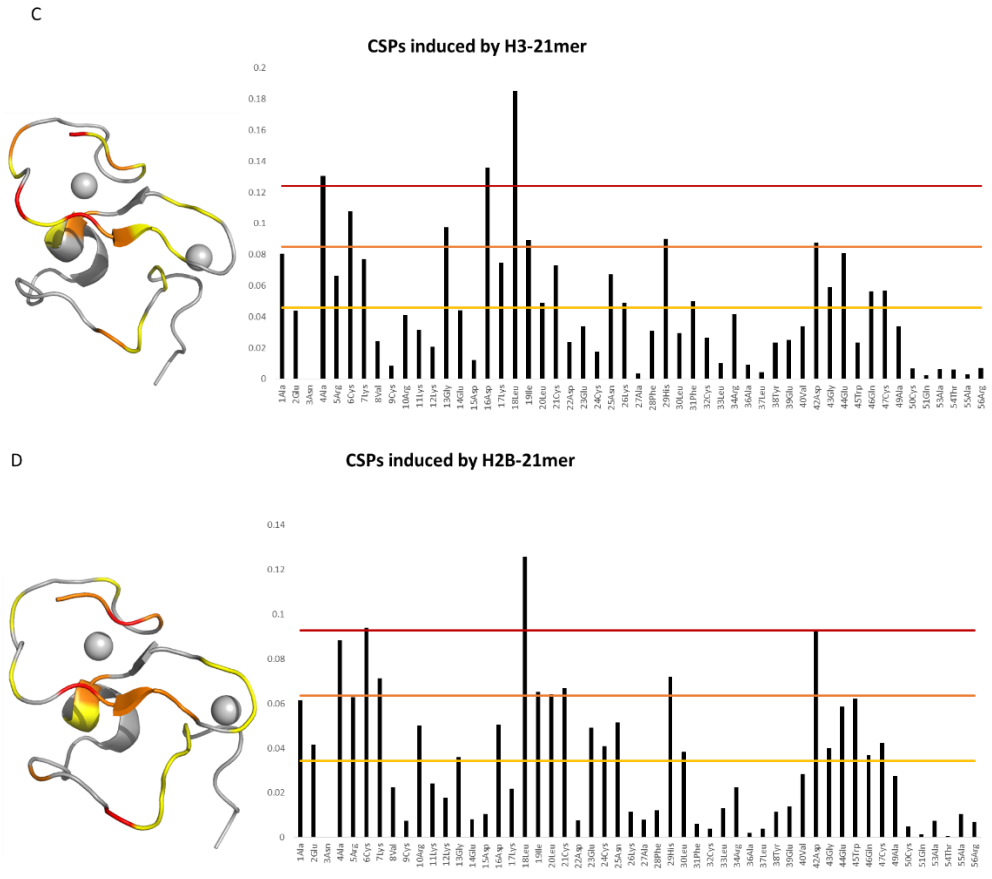


Figure 4.9. Characterization of the interaction between H2A N-terminal tail and PHD of BAZ1B in solution by NMR. CSPs induced by 20-mer histone peptides projected onto NMR structure of BAZ1B (PDB:1F62). CSPs are reported in ppm and grouped according to the intensity of shift. (A-D) CSPs induced by the H4, H2A, H3 and H2B peptides.

Figure 4.10. Sequence alignment of the 20-mer N-terminal H2A and H4 histone peptides. Alignment was performed with Jalview software [170]. Residues 100% conserved are highlighted in blue.

The CSPs map of BAZ1B were compared to the CSPs map obtained for BAZ2A/B PHD. Although the intensities observed were less intense compared to the homologous protein, the hotspots of interaction were conserved (figure 4.11).

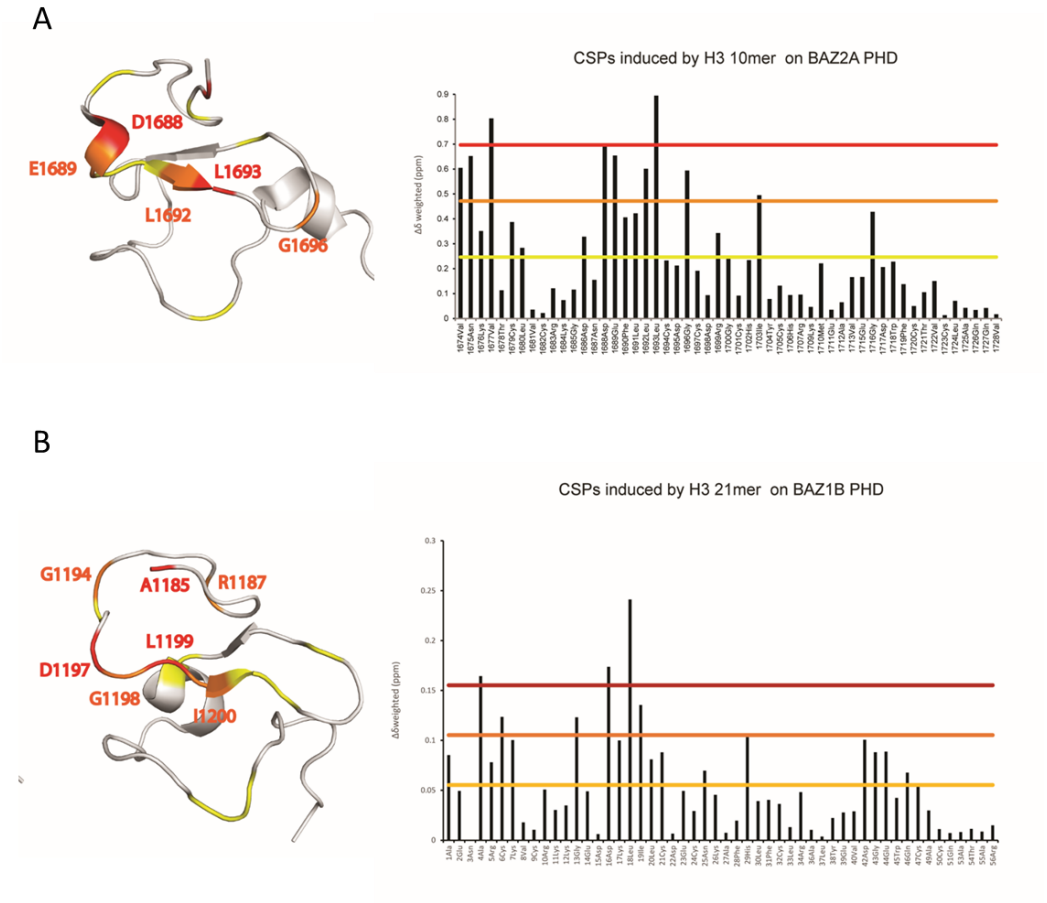


Figure 4.11. CSPs induced by histone peptides on BAZ2A PHD and BAZ1B PHD.

(A) CSPs induced by titration of H3 10-mer on BAZ2A PHD and projected of the structure of BAZ2A PHD (PDB: 4QF2). (B) CSPs induced by titration of H4 21-mer on BAZ1B PHD and projected of the structure of BAZ1B PHD (PDB: 1F62).The residues involved in the strongest and medium shifts are respectively coloured in red and orange. In yellow weak shifts and grey all the other residues with no significant shifts.

To further investigate and validate the reliability of these interactions, isothermal titration calorimetry (ITC) experiments were performed. H2A and H4 peptides were titrated against PHD of BAZ1B. H2A and H4 peptide were tested against PHD of BAZ1B at two different concentrations. Initially at 1.5 mM peptide versus 60-100 uM protein but since the signal detected was too low, it was decided to increase the peptide concentration up to 3 mM for H2A and 2.5 mM for H4 versus 100 uM protein; in this case the buffer concentration was increased up to 100 mM HEPES in order to avoid buffer mismatch due to changes in pH. The only visible result observed was the change from exothermic to endothermic reaction but no binding was detected. Also reduction of the temperature of the experiment did not improve the heat of binding.

4.2.1.8 Histone array screening

To further explore the possibilities of binding of this PHD towards histone peptides, it was decided to use a modified histone array to screen for modifications on the N-terminal histone tails of H3, H4, H2A and H2B that could influence the binding.

Prior to testing the PHD of BAZ1B, a positive control was tested using the 6His -G9a tudor domain, following the manufacturer instructions (figure 4.12 A-B). The amount of protein suggested for the assay is in the range between 10 nm to 1 µM. In our case, the choice of the concentration of BAZ1B to use was directed from previous considerations on the HSQC that showed very weak affinities. It was chosen to use 400 nM, an intermediate value, in order to catch also low affinities without generating high background. From the array, no interactions were spotted. The array appears dominated by a strong background, with the appearance of few spots showing no reproducibility within the duplicates, suggesting no specificity of binding for any histone peptide (figure 4.12 C-D).

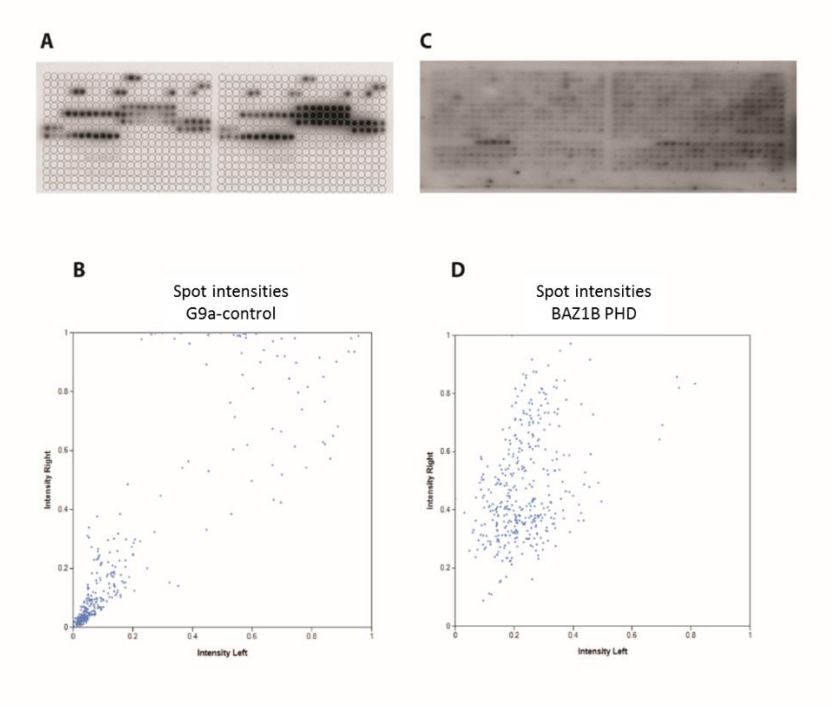


Figure 4.12 A-D. Histone peptide array.

(A) and (C) Images of the ECL detection of the incubation of the histone array duplicates, respectively, with the positive control G9a and with BAZ1B PHD. (B) and (D) Distribution of the spots intensities within the histone array duplicates. In (B) results relative to G9a control and in (D) relative to PHD of BAZ1B.

4.2.1.9. PHD zinc finger of BAZ1B: does it bind any histone tail?

The PHD zinc finger is known to preferentially recognise the histone tails of the nucleosome. Despite this, there are few examples in literature, reporting that this domains could also recognise DNA [173] or have a mere function of structural support to the adjacent domain [174], without recognising any histone tail. Indeed, PHD zinc finger are often found in proximity of another reader domain, for example a BrD, as in the case of BAZ1B. Interesting is the example reported from Zeng *et al.* [175], describing the case of PHD of TRIM28 that does not recognize any histone tail but is essential for the structural and functional integrity of the tandem domain, as well as the case of the PHD of the CBP protein [174].

Based on the unsuccessful attempts at characterizing histone tail binding, and on the observations above, it was decided to expand the structural and biophysical studies by including the adjacent bromodomain in our protein constructs. It was hoped that this could help to understand if the two domains within the PHD-BrD tandem module of BAZ1B may play a role within each other.

4.2.2 Bromodomain and tandem constructs of BAZ1B

4.2.2.1 Bromodomain of BAZ1B

Sequence alignment of BAZ family bromodomains showed that the acetyllysine binding motif is conserved (figure 4.13) [45, 136]. In the literature, it is reported that the bromodomain of BAZ1B binds in a promiscuous way all the acetyllysine on H3, without showing preference for any mark [37].

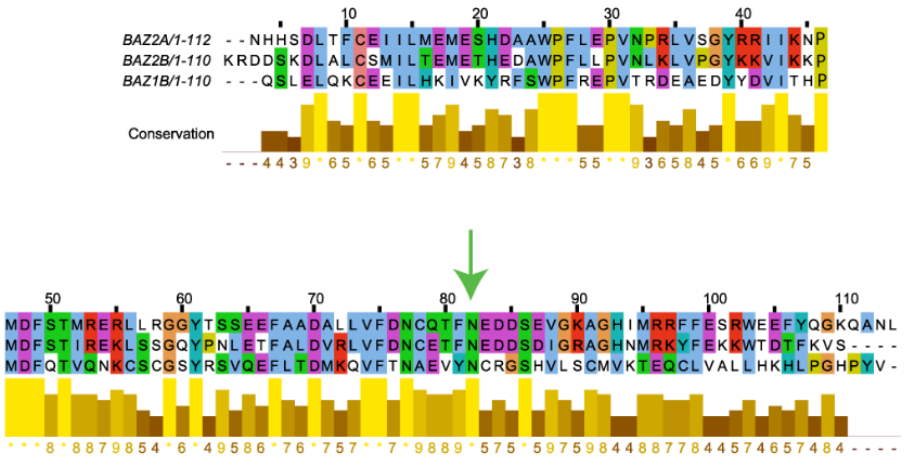


Figure 4.13. BAZ bromodomains alignment.

Sequence alignment of bromodomains of BAZ family was performed using Jalview Software [170]. Colors legend: in blue hydrophobic residues; in red positively charged residues; in magenta, negatively charged residues; in green polar residues; cysteine in pink; glycine in orange; proline in yellow and in cyan aromatic residues. The green arrow highlights the asparagine residue involved in the recognition of Kac. The conservation graph on the bottom is representative of the level of conservation of the residues. 10 is the highest score and is reported as '*'.

In order to verify this hypothesis, expression and purification of the bromodomain of BAZ1B was attempted. Constructs of different length were cloned into pET28a plasmid. The His tag was chosen as purification tag since the His-SUMO tag has a molecular weight comparable with the one of the BrD. I first attempted the purification of a short construct BrD(residues 1343-1444) including only the residues strictly predicted to be part of the bromodomain organization, in order to create a construct prone to crystallization. Phyre2 server was used for the structural prediction of the domain [176], while the IUPred server [177, 178] confirmed that this region is highly structured.

A first expression trial was performed using the BL21 *E.coli* strain and induction of the protein with 0.5mM [IPTG] at two different OD_{600} ~1 and ~ 2. After induction, cells were incubated overnight at 20°C and then spun down, re-suspended in lysis buffer (20 mM Tris-HCl, 5mM Imidazole, 2 mM TCEP, 500 mM NaCl pH 8) and lysed with the French Press. Three fractions were analysed by SDS-PAGE: the total, which includes the cell suspension after cell lysis, the soluble and insoluble fractions, obtained after spinning down the lysate. Results of this first expression trial are reported in figure 4.14.

Overexpressed protein runs on the expected position on the lane, according to its MW (~ 12.7 KDa) but it was found only in the total and insoluble fractions.

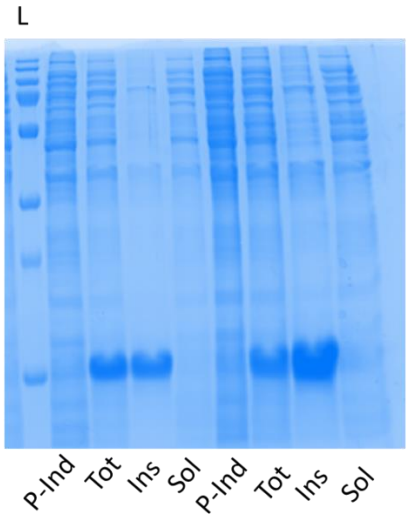


Figure 4.14. Expression trial at two different OD_{600} of induction for BrD(1343-1444) of BAZ1B.

Expression trial of BrD(1343-1444) of Baz1B after induction at OD_{600} ~1 on the left and at OD_{600} ~ 2 on the right. "P-Ind" stands for "pre-induction fraction". "Tot" is the fraction of the lysate straight after lysis. "Ins" represents the insoluble fraction. "Sol" stands for the soluble fraction.

In an attempt to improve the solubility of the construct, BAZ1B BrD was co-expressed with the chaperon GroEL. This approach has been widely used in the last few years to increase folding and solubility of recombinant proteins, and reducing their aggregation. In addition to GroEL co-expression several changes were introduced into the expression and lysis process to maximize the protein solubility. Following induction of protein expression with 0.5 mM of IPTG the cell culture was incubated for 16 h at a lower temperature 16 °C. Furthermore, the lysis buffer was supplemented with 1 M NaCl (to help the solubility of the protein) and 10% glycerol (to stabilize and prevent protein aggregation) [179]. Finally, cells lysis was performed with a milder process through sonication. Unfortunately, also this optimized expression and lysis process did not yield soluble BAZ1B Brd.

Next, the addition of surfactants and a denaturing agent was tested to improve the solubility of this construct. Tween 20 and Triton X-100 were used as detergent

Mild detergents can assist the protein solubility without disrupting native interaction but interacting with hydrophobic portion of the protein sustaining its miscibility in water. Urea was used as denaturing agent. Urea is known to destabilize the protein conformation through direct interactions with polar residues and peptide backbone but also altering the water structure and dynamic [180]. Therefore, the elimination of urea, should lead to the refolding of the protein, in presence of stabilizing agent. This approach is preferentially chosen in those cases where it is possible to check the quality of the refolding process of the protein, for example, testing the activity of the protein or its binding with a known partner. Otherwise, the scepticism on the use of this method is mostly due to the possibility of generating a refolded non-functional protein.

Protein was next expressed in BL21 strains and lysed using a mild method with the lysozyme. As protocol, I followed the one reported by EMBL website (www.embl.de). 1 mg/mL of lysozyme was added to the lysis buffer with 5 mM EDTA to enhance the activity of the lysozyme. Cell were incubated for 2h at 4°C under agitation. Subsequently, the addition of 6 M urea, 0.2% Tween 20 and 0.2% Triton X-100 was tested but no improvement of solubility was observed.

To improve the solubility of the protein, I designed two other constructs expanding the number of amino acids at the N- and C-terminus. IUPred software [177, 178], which estimates the disorder tendency in proteins, helped to decide which region to choose to increase the solubility of the construct. The 6His-SUMO tag, which is known to increase the solubility of the constructs, could not be used since its MW is comparable with the one of the bromodomain and this would make more difficult to check the quality of the samples by SDS page gel.

BrD(1321-1449) and BrD(1310-1449) were obtained by cloning strategy using the same plasmid as above.

The expression of these two constructs was tested in two different *E.coli* strains, BL21(DE3) and Rosetta. Induction was performed at OD_{600} 0.8 and with 0.5 mM IPTG. Cells were incubated overnight at 16°C. After this, cells were spun down and pellet was re-dissolved in 20 mM Hepes at pH 8.5, 1 M NaCl, 1 mM DTT supplemented with 10% glycerol. The calculated pI for BrD(1310-1449) is 7.2 (Protparam calculation [181]), this motivated to use a buffer at pH 8.5. Results of the expression trial are reported in figure 4.15.

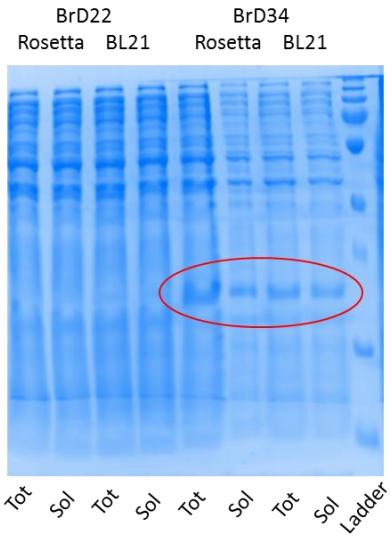


Figure 4.15. Expression trial for the longer constructs of BAZ1B Bromodomain.

Expression trial of construct BrD(1321-1449) on the left and BrD(1310-1449), on the right. Total (tot) and soluble (sol) are reported. Circled in red the potential bands of BrD(1310-1449) (MW 17.4 KDa). No spotless bands for BRD22.

For BrD(1321-1449) no spotless bands are detected in the expected MW; for BrD(1310-1449), a band that may correspond to the MW of our construct could be observed, that was present in both total and soluble fraction with the same intensity (red circle). However, the level of expression is very low in these conditions since is comparable to the background protein of the strain used. In order to verify if the band highlighted is really relative to the bromodomain construct of BrD(1310-1449), scaling up of the expression is needed with further optimization of purification conditions.

4.2.2.2. Expression trial of the PHD-BrD tandem modules of BAZ1B

Following these trials for the individual BrD of BAZ1B, it was decided to move to express also the full tandem module PHD-BrD.

A PHD –BrD construct of BAZ1B(1321-1449) was sub-cloned in pCRi11b plasmid in order to express the protein as 6His-SUMO tagged protein. In this case, the different MW between protein and affinity tag allows to differentiate them on the SDS page.

Expressing this construct under identical conditions employed for the other PHD-BrD tandem constructs did not lead to soluble protein. It was therefore decided to undertake

a more systematic approach. After initial incubation at 37°C, the induction was induced with two different IPTG concentrations: 0.1 mM and 1 mM. Overnight incubation was performed at three different temperatures: 37°C, 25°C and 16°C. Expression was checked at different time: after 2h, after 4h and after overnight incubation. Following cell lysis by sonication, fractions representative of different conditions were loaded on a SDS page gel to induction of protein expression. As reported in picture 4.16, the lanes are all similar to the pre-Induction lanes and no overexpressed band is showing in any of them at the expected MW (~43 KDa).

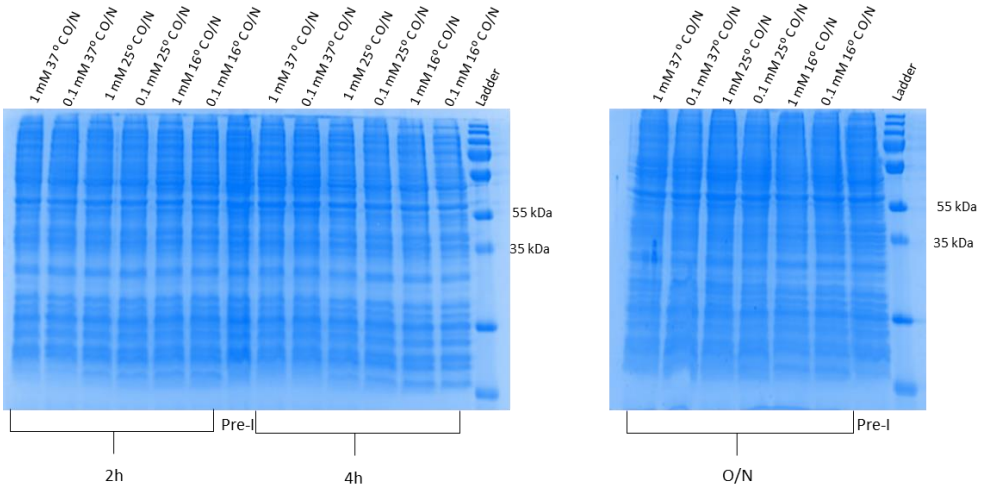


Figure 4.16. SDS-page gel of expression trial for PHD-BrD BAZ1B.

Expression trial of the tandem construct of BAZ1B in BL21 *E.coli*. Each lane is labelled with the condition of expression. No bands are detected in the expected region at ~43 KDa.

This test was performed using BL21(DE3) *E.coli* strain. Further trials were carried out using the Rosetta strain and, more importantly, C41(DE3) [182] strain that is known to help the expression in case of toxic proteins that are easily degraded in other strains. Also in this case, no overexpression of the target protein was observed. Sequencing analysis confirmed that the region of the gene after the T7 promoter is preserved, so the lack of expression could not be related to an issue with the DNA quality of the plasmid.

In summary, it has proven challenging to identify and characterize the structure and a putative binding partner of BAZ1B PHD. It also proved so far to be mostly unsuccessful to express soluble protein constructs of BAZ1B tandem and BrD domains. Due to time

constraints, it was decided during my PhD to redirect efforts in other directions. In light of the results obtained, future plans should consider alternative expression systems as, for example, insect cell that are useful when protein cannot be expressed in *E.coli* cause a better folding or crucial post translational modification are needed. Xiao *et al*. were the only group reporting expression of the PHD-BrD tandem of BAZ1B and, specifically, in insect cells, although still yielding a very low amount of soluble protein [105].

4.3. TANDEM CONSTRUCT OF TRIM66

In parallel to the study on the elucidation of the histone binding partner for BAZ1B tandem modules, I investigated on the PHD-BrD tandem module of the TRIM66 protein. In this protein, differently from BAZ1B, the two reader domains within the tandem were located very close to each other likely suggesting cooperativity or interactions between them, in the recognition mode of a potential binding partner. It was therefore decided to express TRIM66 as a tandem PHD-BrD module.

Four different PHD-BrD tandem constructs of TRIM66 were subcloned in the expression vector pET28a, inserting the TEV protease cleavage site, and expressed as hexahistidine tagged protein. Amongst the four constructs, the one that showed higher solubility was selected for the biophysical assays, T66(966-1159). Its gel filtration chromatogram is reported in figure 4.17. The elution is observed at the expected volume, according to its MW ~22.5 KDa.

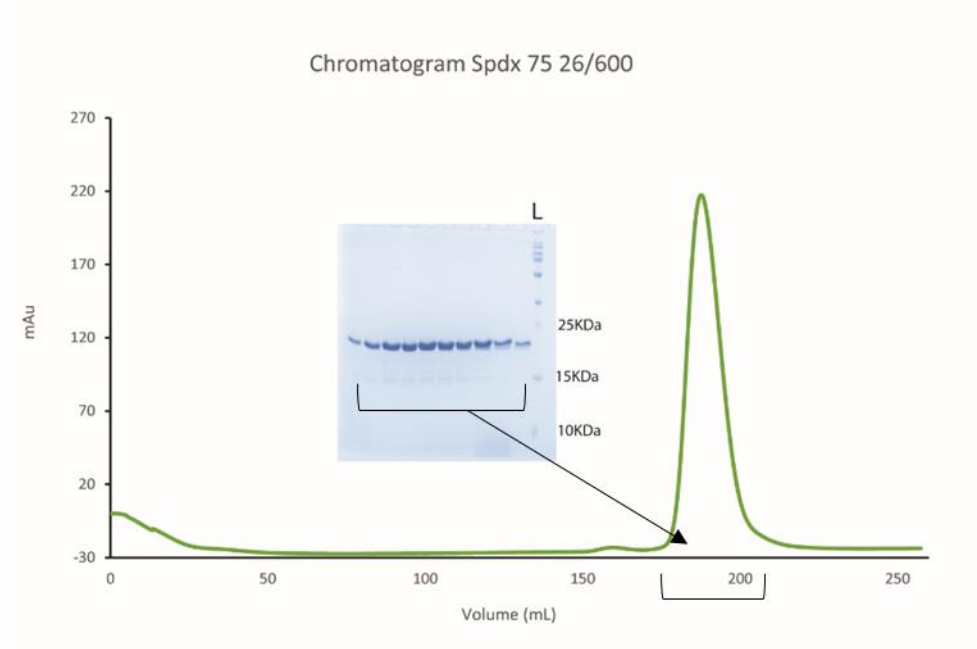


Figure 4.17. Gel filtration chromatogram of TRIM66 PHD-BrD and SDS-page gel.

Chromatogram of the gel filtration of TRIM66 PHD-BrD in green. On y axis, absorbance at 280 in mAu. On x axis, volume in mL. The column used for the gel filtration was a Superdex 75 26/600 (GE Healthcare). 16% SDS-page gel showing the bands of TRIM6, in the expected region (MW~22.5 KDa).

Next, a DSF screening was performed to identify the buffer conditions that provide the highest stabilization of the protein. Most suitable conditions were found in a pH range between 7.0 and 8.5 in Hepes, Tris and phosphate buffer.

4.3.1 TRIM66 PHD-Bromo binds amino terminal tail H3

4.3.1.1 PHD zinc finger

The sequence alignment of the PHD domains of the TRIM family proteins highlighted that, the histone binding region is highly conserved for TRIM24, TRIM33 and TRIM66 (red circles in figure 4.18), with the exception of TRIM28 which is known it does not recognize any histone tail [175]. In addition, the aromatic residue that is signature of specific methyllysine recognition at position 9 (K9me) is conserved (green arrow in figure 4.18), while the aromatic cage typical of the H3K4me3 reader is not present. PHD of TRIM24 and TRIM33 were already known to recognize the H3 histone tail [51, 62], herein the alignment suggested that also the PHD of TRIM66 might be capable of binding to the H3 tail.

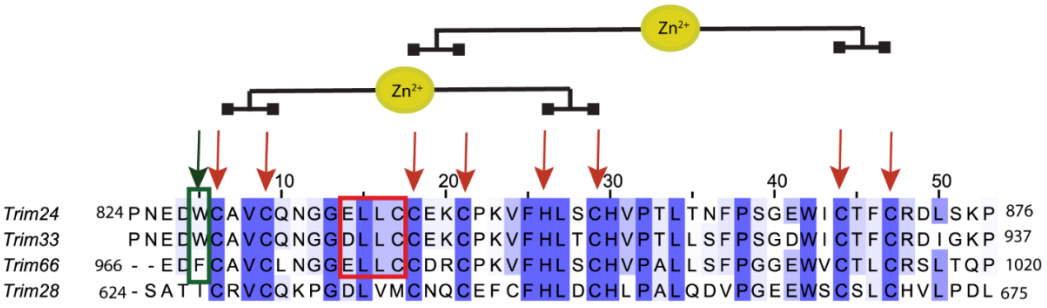


Figure 4.18. Alignment of the PHD zinc finger domain within the TRIM family.

Sequence alignment made by Jalview [170] of the PHD fingers within TRIM family**.** Cys and His of the conserved motif of the PHD zinc finger are highlighted by the red arrows. The conserved H3 histone-binding motif is circled with the red box. The green arrow and green box highlight the aromatic residues conserved through TRIM33, TRIM24 and TRIM66. In blue residues 100% conserved through the TIF1 family.

To test our hypothesis of binding based on the sequence alignment, a 10-mer peptide resembling the unmodified N-terminal tail of H3 (ARTKQTARKS) was synthetized and tested by ITC. Titration confirmed binding of the PHD-BrD of TRIM66 to the H3 histone tail reporting a K_D of 6 µM. Modifications on K4 and K9 were tested and showed a decrease of binding that in the case of K9 was less pronounced than in K4 (figure 4.19). Indeed, acetylation of K4 led to K_D of binding affinity of 130 µM while acetylation on K9 moderately affected binding, increasing K_D up to 10 µM. I decided to test also the effect of methylation on both K4 and K9. 10-mer peptides were synthetized with both mono and double methylation on K4 and K9. It was not possible to synthesize the peptide containing try-methylated lysine since the required amino acid could not be purchased at the time of the experiment. Summary of the thermodynamic data is reported in Table 4.1. ITC titrations showed how methylation on K4 sensibly reduced the affinity of binding (table 4.1). In contrast, double methylation at position K9 was tolerated (table 4.1).

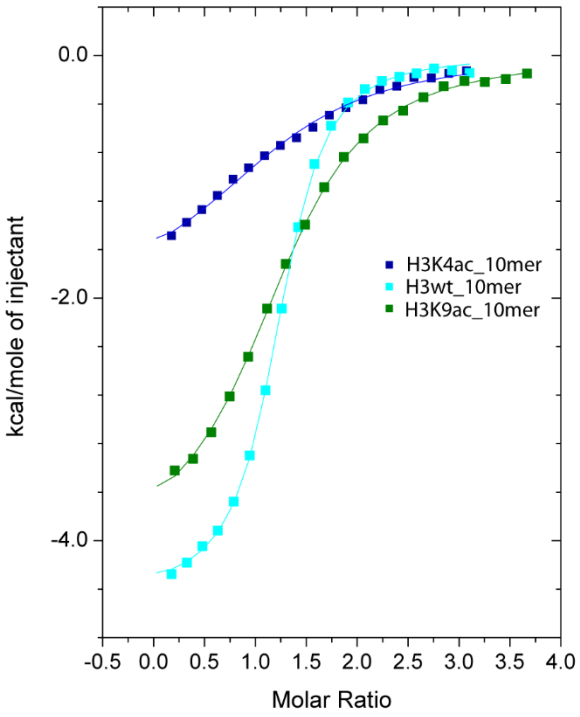


Figure 4.19. Raw ITC-binding curves of the H3 10-mer peptides titrated into PHD-BrD of TRIM66.

Peptide	K_D (μM)	N	ΔH (kcal/mol)	TΔS (kcal/mol)	ΔG (kcal/mol)
H3 wt 10-mer	5.8 ± 0.6	1.3 ± 0.1	-4.3 ± 0.7	2.9 ± 0.6	-7.2 ± 0.1
H3K4ac 10-mer	131 ± 19	1.3 ± 0.1	-4.0 ± 0.4	1.3 ± 0.5	-5.3 ± 0.1
H3K9ac 10-mer	10.7 ± 0.9	1.3 ± 0.0	-4.0 ± 0.1	2.7 ± 0.1	-6.9 ± 0.1
H3 K4me 10-mer	23 ± 1.1	1.7 ± 0.01	-3.06 ± 0.01	3.26 ± 0.01	-6.32 ± 0.01
H3K4me2 10-mer¥	31.3 ± 0.2	1.32 ± 0.03	-2.63 ± 0.07	3.5 ± 0.1	-6.14 ± 0.04
H3K9me2 10-mer	6.5 ± 0.1	1.5 ±0.1	-4.0 ± 0.01	3.08 ± 0.01	-7.08 ± 0.01
H3 wt 21-mer	10.2 ± 1.2	1.56 ±0.02	-4.13 ± 0.03	2.68 ± 0.03	-6.81 ± 0.01
H3K9ac 21-mer¥	5.6 ± 0.3	1.50 ± 0.01	-3.83 ± 0.03	3.33 ± 0.05	-7.2 ± 0.1
H3K14ac 21-mer¥	5.1 ± 0.5	1.52 ± 0.02	-3.06 ± 0.05	4.1 ± 0.1	-7.2 ± 0.1
H3K18ac 21-mer	2.9 ± 0.5	1.7 ± 0.1	-6.5 ± 0.1	1.1 ± 0.1	-7.5 ± 0.1
H3K14acK18ac 21-mer	2.6 ± 0.3	1.3 ±0.1	-8.5 ± 0.4	-0.8 ± 0.3	-7.6 ± 0.1
H3K23ac 30-mer	0.72 ± 0.01	1.31 ± 0.01	-4.67 ± 0.02	3.71 ± 0.02	-8.38 ± 0.01
H3K27ac 30-mer	1.2 ± 0.1	1.61 ± 0.02	-3.35 ± 0.02	4.72 ± 0.04	-8.06 ± 0.02

Table 4.1. Summary of the ITC-based thermodynamic binding parameters for complex formation between TRIM66 PHD-BrD and different H3 peptides.

Titrations were performed at 25 ^{o}C in triplicate or duplicate, except where differently specified, and values reported are the means ± s.e.m.

¥ = titrations were performed only once and the errors are calculated by MicroCal ORIGIN software.

A model of the structure prediction of Trim66 was generated using Rosetta server [183] and aligned in Pymol with the PDB structure of Trim33 in complex with H3K9me3 (PDB: 3U5O) (figure 4.20 and 4.21). Structure of TRIM33 in complex with H3 histone showed that H3K4 is involved in hydrogen bond with TRIM33 through the side chain of D885 and D889 plus the carbonyl of N887, at the N-terminus of the domain. From the

superposition shown in figure 4.20, Trim66 is predicted to preserve the same set of binding with the conserved N969 and D971 while the D885 of Trim33 is replaced by I967 in Trim66. The presence of this hydrophobic residue could justify why additional methylation on K4 are not completely detrimental and the binding is still retained, although with lower affinity (K_D 23 µM for K4me and K_D 31 µM for K4me2). Instead, the presence of an acetylation mark, which neutralizes the charge on the lysine side chain, showed a remarkable loss of binding (K_D 130 µM).

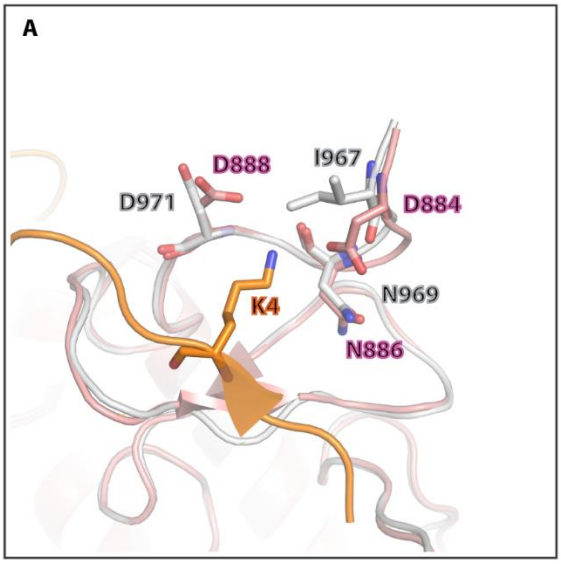


Figure 4.20. Superposition of TRIM33 and prediction model of TRIM66.

Zoom in on the interaction between the K4 of the histone peptide and the superposition of the predicted model of TRIM66 (white cartoon and sticks) with the crystal structure of TRIM33 (pink cartoon and sticks) in complex with H3K9me3K14acK18ac peptide (cartoon and orange sticks). Each residue is labelled according to the colour of the protein which it belongs.

Hypothesis on the mode of binding of H3K9me2 towards TRIM66 PHD-BrD were using the same superposition with the Rosetta model of TRIM66. As reported in the alignment in Figure 4.18, the aromatic cage required for the methyllysine binding is not conserved in the TRIM PHDs but the presence of W889 in TRIM33 is able to coordinate the K9me3 on H3 through the cation-π recognition. In TRIM66 the aromatic residue is replaced by F972 and the Q894 in TRIM33 is substituted by V977 that can assist the coordination of the methyllysine (figure 4.21). A study published by Gallivan and Dougherty [54] suggested that the Trp is more suited to form a cation-π recognition than Phe or Tyr. This may

account for the lower binding affinity showed from TRIM66 than TRIM33 toward the H3K9me2 histone. Taking in account these observations, it would be also interesting to investigate potential methylation occurring on the R8 and how can affect the interaction to the PHD, since methylated arginine are more favoured in the cation-π recognition than Lys.

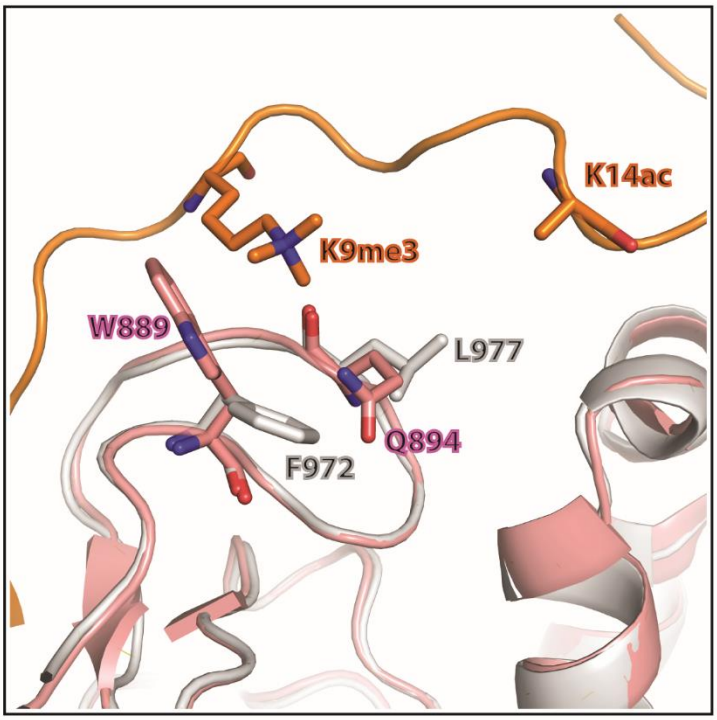


Figure 4.21. Superposition of TRIM33 and prediction model of TRIM66.

Zoom in on the region of binding of H3K9me3 and TRIM33 (cartoon and stick in pink). W889 involved in the cation-π recognition is drawn as stick. Superposition of the complex TRIM33 (pink cartoon) with H3K9me3 peptide (orange) and the predicted model of TRIM66 (grey cartoon).

4.3.1.2 Bromodomain

Sequence based alignment of the bromodomains within the TRIM family implied that also TRIM66 BrD is an acetyllysine reader since the F/N residues in the binding of the Kac are conserved (green arrow in figure 4.22).

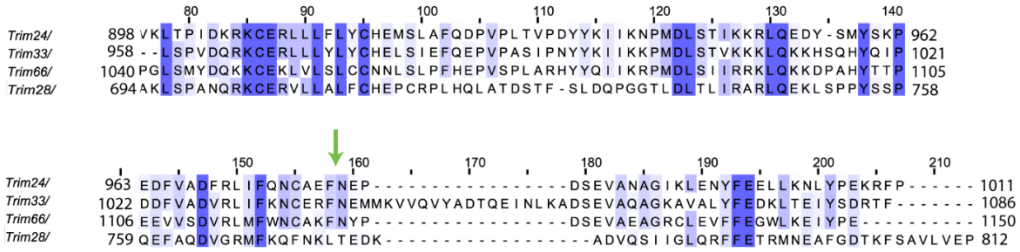


Figure 4.22. Sequence alignment of bromodomain of TRIM family.

Sequence alignment of the bromodomain of TRIM family. Highlighted in blue are the residues 100% conserved. The F/N residues known to recognize Kac are conserved only for TRIM24, TRIM33 and TRIM66, while TRIM28 does not show any conservation. Indeed, TRIM28 does not recognize any Kac [175].

In first instance, 10-mer peptides were synthesized, containing a single Kac, to identify any specificity of binding. Tested peptides were containing K9ac, K14ac, K18ac, K23ac, K27ac and K36ac in the middle of the sequence, expecting that the BrD would have engaged a unique acetyllysine residue. No significant binding was detected with any of these peptides. Therefore, I decided to synthesize longer peptides carrying the same Kac, in order to investigate for any potential cooperative effect. H3K9ac 21-mer and H3K14ac 21-mer did not show any change of binding when compared to the H3 21-mer wt (Table 4.1 and figure 4.23) and, more importantly no relevant variation in the heat signals were observed. H3K18ac 21-mer reduced the K_D to 3 µM compared to the unmodified peptide (K_D ~ 10.2 µM) with an associated ΔH of -6.5 kcal/mol (figure 4.23). This increment of the heat signal could account for a cooperative effect during the binding of the two reader domains to the double marks on the peptide: the unmodified region from the PHD and the K18ac from the BrD. Double acetylation H3K14acK18ac did not improve the binding affinity compared with the single acetylation on K18ac, although a further increase in absolute value of ΔH was observed (ΔH ~-8.5 kcal/mol, figure 4.23 and Table 4.1). These data would suggest that the binding is mainly driven by the acetylation on K18, while K14ac alone seems do not play any relevant role, but in combination with K18ac could mediate cooperativity of binding. It could be speculated that, admitting a linear binding mode of the unmodified H3 peptide to the PHD of TRIM66, it might be that the K14ac is located too far from the BrD pocket of the same molecule of TRIM66 for being able to bind it (Figure 4.21). Analogous situation was

reported for TRIM33 [51]. Remarkable increment in affinity were detected upon titration of TRIM66 with 30-mer peptides carrying respectively K23ac and K27ac, with K_D decreasing to 0.6 µM and 1.1 µM, respectively. For these peptides no changes in ΔH were observed compared with shorter and unmodified peptides (Table 4.1 and figure 4.23).

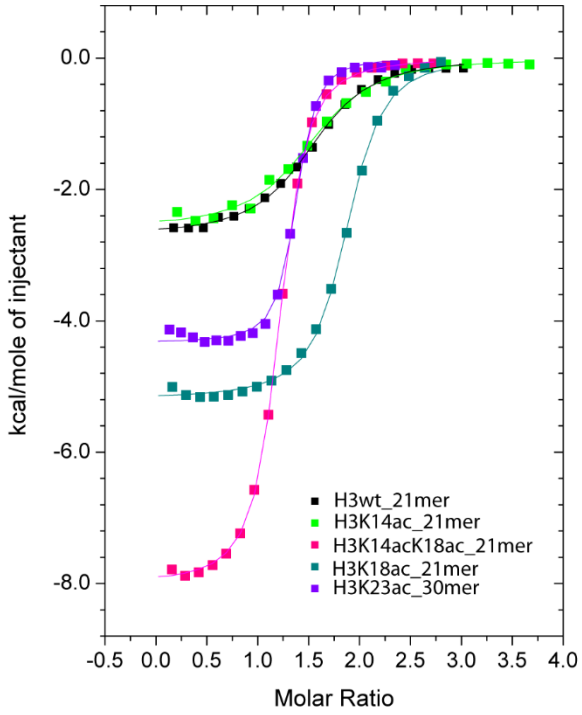


Figure 4.23. Raw ITC-binding curves of the H3 21-mer and 30-mer peptides titrated into PHD-BrD of TRIM66.

4.3.2 Crystallization efforts to gain molecular insights of histone recognition

To understand the molecular basis of H3 recognition by the PHD-BrD of TRIM66, I wanted to solve the structure of this tandem module either in the *apo* form and/or in complex with the cognate histone. A crystal structure of the *apo* form might have been useful to perform soaking experiments in the case the histone pockets of PHD and bromodomain were free from crystal contacts. Crystallization screening of sparse matrix and grid screen were set up with two different constructs of TRIM66 tandem at different concentration of protein in a range between 5 and 24 mg/mL. Crystallization drops were

visually inspected and most of them resulted in precipitate or microcrystals formation. Few drops that showed interesting formation were attempted to be optimized.
For example, in the Morpheus screening two similar conditions were identified that led to microcrystals (figure 4.24).

A7 Morpheus screen T=20°C
10% w/v PEG4000
20% v/v glycerol
0.03 M $CaCl_2$
0.03 M $MgCl_2$
0.1 MOPS/HEPES-Na pH 7.5

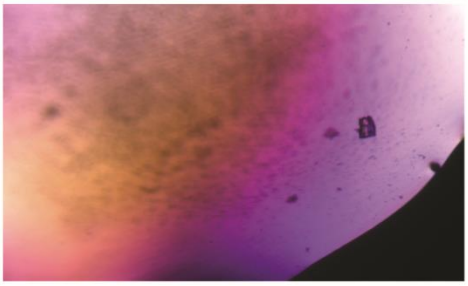
A10 Morpheus screen T=20°C
10% w/v PEG8000
20% v/v ethylene glycol
0.03 M $CaCl_2$
0.03 M $MgCl_2$
0.1 Bicine/Trizma pH 8.5

Figure 4.24. High magnification pictures of crystallization drops.

Crystallization drops from Morpheus screening reporting microcrystals from *apo* TRIM66 PHD-BrD at 20°C. On the bottom of the figures details of the crystallization conditions.

Conditions within the two drops have high similarity of components. In both cases the pH is in the range found as the most stable for the protein and far from the theoretical pI (~5.8). Low MW PEGs were used as precipitant, which increase the molecular crowding inducing separation of the macromolecules from the solution, in combination with 20% of an organic solvent (glycerol or ethylene glycol) that reduces the dielectric constant of the solvating medium. In this context, the presence of divalent salts helped the protein to avoid copious precipitation keeping it in the salting-in region. In order to push the conditions towards the ideal conditions for crystallizations, few parameters were changed. Different concentrations of protein were explored between 10 and 4 mg/mL simultaneously to different concentration of the divalent salts.
Also the effects of pH and protein concentration were tested. Four different pH were explored (8.1, 8.3, 8.5 and 8.7). One of these conditions led to the formation of a drop

containing a tiny crystal, still too small to be harvested and shot on the X-ray generator (figure 4.25).



10% w/v PEG8000
20%v/v ethylene glycol
0.03 M $CaCl_2$
0.03 M $MgCl_2$
0.1 M Bicine/Trizma pH 8.3

10.6 mg/mL Trim66

Figure 4.25. Magnified view of the tiny crystal obtained from optimization. Conditions are described on the right of the figure. Protein concentration used of TRIM66 PHD-BrD was 10.6 mg/mL.

Also the effect of the concentration of the PEG8000 and ethylene glycol was explored, keeping constants all the other conditions and the protein concentration. This led to the generation of few drops with tiny crystals, bigger than microcrystals, but still too difficult to be harvested (figure 4.26).

5% w/v PEG8000
15% v/v ethylene glycol

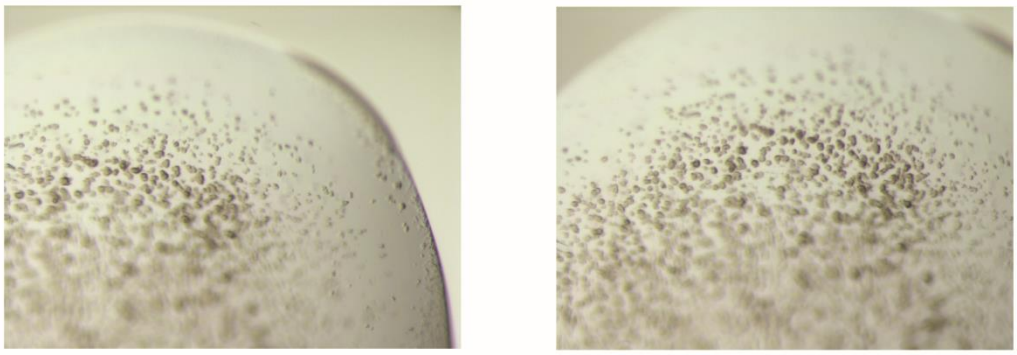
5% w/v PEG8000
25% v/v ethylene glycol

Figure 4.26. Magnified view of the tiny crystal obtained from optimization.

Microcrystals obtained upon optimization at room temperature of crystallization condition for TRIM66 PHD-BrD apo form. Conditions used are detailed right below the picture.

Crystallization plates were set up also with the tandem module in complex with H3 wild-type 10-mer peptide (protein/peptide ratio 1:3), withH3K14ac 21-mer peptide (protein/peptide ratio1:1.5 and 1:2) and H3K18ac 21-mer at different protein/peptide

ratio. Figure 4.27 illustrates some of the hits identified for TRIM66-H3 10-mer peptide complex in JCSG+ screen.

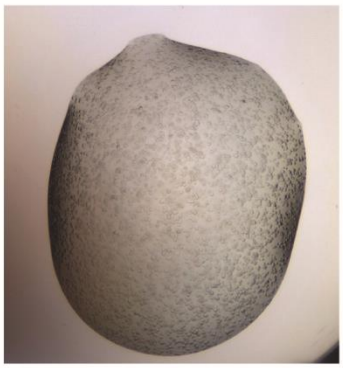


G2 JCSG+
100 mM HEPEs pH 7.5
20 mM MgCl2
22% Poly(acrylic acid sodium salt) 5100

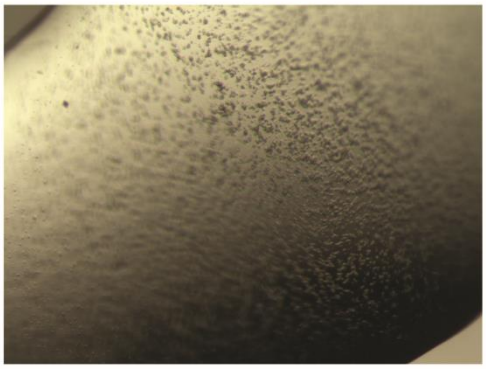
Figure 4.27. Magnified view of the microcrystal or quasi crystals obtained from G2 in JCSG+. Conditions used are described below the picture.

In the scale up to larger drops in the optimization grid the amount of protein was reduce to 8mg/mL from the initial 14.6 mg/mL of the small drop of the screening. This is needed to compensate the loss of protein that is generally occurring in the small drops since in small drops at the interface drop/air lot of protein is lost so the real concentration is lower than the theoretical one. Optimization tested different precipitant concentrations between w/v 8% and 22% at pH of 7.0, 7.2, 7.5 and 7.8, but only precipitation was observed.

Interesting formation resembling quasi crystals or small crystals were observed for the complex of the tandem with H3K14ac 21-mer peptide in the JCSG+ plate (condition H1, figure 4.23). In this condition the pH 5.5 was found very close to the calculated pI. As precipitant there was only magnesium formate. The influence of the temperature seems to be relevant in increasing the size of the formations (figure 4.28).



T 4°C
100 mM Bis-Tris pH 5.5
300 mM $Mg(HCOO)_2$



T 20°C
100 mM Bis-Tris pH 5.5
300 mM $Mg(HCOO)_2$

Figure 4.28. Magnified view of the microcrystal or quasi crystals obtained from H1 in JCSG+.

Microcrystals and quasi crystals observed at 4°C (left) and room temperature (right) for the complex PHD-BrD TRIM66 with H3 10-mer peptide.

In the optimization process different conditions were explored to expand the range of the precipitant concentration between 100 and 600 mM, keeping the pH constant. It was observed that microcrystals were bigger at increasing concentration of precipitant but still too small to be harvested on a loop. Keeping constant the protein/peptide ratio at 1:3, three different protein concentrations were used: 9mg/ml, 15 mg/mL and 20 mg/mL. Different pH conditions were tested in presence of low precipitant concentration (0.2 mM or 0.3 mM) and with the addition of 5% glycerol and 8- 20% w/v PEG3550 to assist the precipitation. Also the effect of addition of higher NaCl (1-2 M) concentration was tested but no crystal formation was observed. No success was gained even testing different drop size and protein/reservoir ratio as well as the addition of Al's oil (1:1-paraffin:silicon oil) which should decrease the rate of the vapor diffusion and consequently the formation of the *nuclei*.

An additive screening was also performed using as starting condition the one in figure 4.23, with the aim to identify a reagent able to perturb sample-sample and sample-solvent interactions in a way to induce crystalline formations. Also a matrix microseeding experiment (MMS) was performed using as seed the quasi crystals in

condition H1 of JCSG+ (figure 4.23). In both cases, only microcrystals formed which would have required further optimization.

Observation of the drops through all the sparse matrix screening, highlighted that microcrystals or quasi crystals formations were mainly present in all the condition with pH close to the calculated pI of the protein (~5.8). In contrast, at pH < 5.5 protein precipitation was consistently observed. Small MW PEGs and $Mg(HCOO)_2$, $MgCl_2$, NH_4AcO, NaCl and $Mg(AcO)_2$ were the favourite overall precipitants. These considerations could lead future works towards the development of a 96-well screening containing all the precipitants listed above plus those with similar properties, in order to explore a more selective range of reagents and their effect on the protein using different pH and concentrations range.

In summary, high-affinity binding partner histone peptides were identified for the PHD-BrD tandem of TRIM66. However, unfortunately, no suitable crystals could be grown for the protein either *apo* or in presence of peptide.

4.4 DISCUSSIONS

In summary, in this chapter, I have presented a preliminary study on the histone recognition of the PHD-BrD tandem module of two chromatin-binding proteins: BAZ1B and TRIM66.

Deorphanization studies performed on the single PHD domain of BAZ1B, using BLI, protein observed NMR and ITC experiments, did not reveal specificity of binding towards any histone peptide tested. The presence of a lysine residue (Lys1998) in BAZ1B PHD domain located in a key hydrophobic patch for histone binding was hypothesized to be responsible for the lack of specificity. For this reason, future work on this reader domain should focus towards the elucidation of the potential influence of this lysine within the histone pocket. To address this point, mutagenesis experiments on both proteins BAZ2A/B and BAZ1B are warranted.

From literature, we have also learnt that some PHDs can have a role of structural support within tandem epigenetic domain unit [175]. To further investigate on the mutual

relationship within the two reader domains, it was decided to express also the bromodomain construct and the tandem construct of BAZ1B. Work done so far has only led to the identification of a potential soluble construct for BAZ1B BrD, but its expression will need to be further optimized. The tandem module did not show any expression in *E. coli*. Therefore, an alternative approach must be undertaken, as for example, the expression either in insect cells, already reported in literature [105], or in mammalian cells. The presence of the long linker (108 AA) between the two domains could be detrimental to the formation of a stable tandem construct and could require further stabilization through the binding to another partner. Indeed this linker, as for the long linker in BAZ2A/B, could act as a scaffold for the recruitment of potential binding partner. Pull-down experiments, using the single domains and the tandem constructs, could help to elucidate preferential binding partners for the PHD-BrD of this chromatin-associated protein.

The protein TRIM66, in contrast to BAZ1B and BAZ2A/B proteins, has the two readers, PHD and bromodomain, located very close to each other, with only a short spacer between them. PHD sequence alignment within the TRIM family allowed formulating hypotheses of binding that were confirmed by ITC experiments. Indeed, TRIM66 recognizes unmodified H3 tail and seems to tolerate methylation in position K9, presumed to be the result of cation-π recognition with residue F972. Studies on the specificity of binding of the bromodomain have shown preference of binding for the acetylation in position K18 and K23 and the binding was favoured in case of the longer peptides. It was observed that different length of the peptides can impact on the binding affinity. For this reason, in order to properly identify whether TRIM66 BrD has preferential mark recognition, the tested peptides should all have comparable lengths.

The results presented provide a preliminary investigation of histone binding recognition to TRIM66, which have nevertheless laid out a direction of investigation to be pursed and raised new questions. It will be interesting to analyse in future the contribution of binding of double marks on the same histone tail, as for example, K9me3 in combination with K18ac or K23ac. Structural insights at atomic level will be crucial to elucidate the binding mode and rationalize the observations of binding from the ITC experiments.

Crystallization results to date have not led to suitable single protein crystal, and further studies could move towards the crystallization of the complex between TRIM66 tandem and the identified peptides that showed higher binding affinity, to ensure full saturation and stability of the protein-peptide complex. If structural information will not be achieved by crystallography, either protein-observed NMR or computational studies could be a valid alternative to gain information into the molecular recognition of TRIM66 and its histone partner.

CHAPTER 5.

5.1 CONCLUSIONS AND FUTURE DIRECTIONS

To date, despite the numerous progress achieved on the understanding of the mechanism beyond the epigenetic regulation, still more need to be accomplished toward a complete overview of epigenetics and how they relate to diseases.

In this context, the intense research in the identification of chemical probes able to tackle specific domains of chromatin-associated proteins could furnish valid tools to interrogate epigenetics and new chemical scaffolds to drug the epigenome.

Currently, important successes have been achieved in the discovery of inhibitors valid to target histone-modifying enzymes and reader domains [70]. Most of these molecules, are largely used as chemical probes to study biology. However, in addition to their use as chemical tools, a few epigenetic inhibitors have already been approved from the FDA as drugs, while others have also successfully entered clinical trials [80] . The majority of the reader domains successfully targeted to date are bromodomains [30, 184] or few methyllysine binders, as MBT and chromodomains [83].

PHD zinc finger reader domains have been linked to several disease states, so they are also emerging as attractive candidates in epigenetic drug discovery. However, so far, only two systematic studies to assess their ligandability have been reported [56, 57]. The paucity of information about the ligandability of these reader domains, which could sets aside important disclosures, motivated this work.

In the first part of my thesis project, I aimed to elucidate the binding mode of the PHD of BAZ2A and BAZ2B towards their histone binding partner and to probe the ligandability of these domains with small fragments. I initially solved the crystal structure of the PHD of BAZ2A in complex with the H3 histone tail. This allowed the identification of the helical conformation of the H3 peptide upon binding with the PHD. Further work performed within the Ciulli Lab helped to define the helicity of the H3 peptide as an essential feature of the histone tail binding mode. This was further confirmed by the structure of the PHD of BAZ2A in complex with the double alanine mutant H3 peptide, which increases the peptide binding affinity. It was further observed that the helical

conformation in H3 was induced by a short helical turn preceding the β-strand of the PHD. This observation combined with biophysical analysis was essential to lead to the identification of a subclass of PHD characterized by an acidic patch on the helical turn, responsible for inducing helicity in the H3 peptide. The structural importance of this acidic patch was corroborated by mutagenesis experiments, which inverted or abolished the charge on the helical turn, inducing a considerable reduction of affinity for the histone tail. Together, these findings from my PhD work and the work of others in the group have contributed to the realization that a subclass of PHD fingers have the ability to recognize the histone tail specifically in helical conformation. This suggests that the secondary structure of the histone tails could constitute a further level of regulation in the epigenetic processes.

In a second part of the project, I investigated the ligandability of the PHD zinc finger of BAZ2A and BAZ2B. Using HSQC NMR, as a main technique for monitoring fragment binding, I performed a biophysical validation of 19 chemical hits identified by *in silico* screening (performed by collaborator Dr. Xavier Lucas within the group). Nine hits were confirmed as potential binders and K_D of binding and LE were provided using CSPs analysis. One hit led to a fragment bound structure with partial electron density. The binding mode of this fragment was used as scaffold to further improve affinity of binding for the fragment. This study identified four hits, all validated by NMR. I solved the structure of the PHD of BAZ2A in complex with one of these optimized fragments and of BAZ2B with two of these optimized fragments.

To the best of our knowledge, these are the first crystal structures of PHD fingers solved in complex with fragments binding to the histone pocket. Comparison of the structures highlighted some differences in the binding mode of the same fragment towards the two different proteins. These observations can be used to guide future design towards expansion of these hits, in a way to enhance potency and provide some selectivity of binding within the same family. Furthermore, fragment optimization could take advantage of the importance of the helicity to increase binding affinity. Indeed, the linkage of these fragments with a synthetic scaffold able to mimicry a helical pattern could be beneficial in future to increase the affinity of the small molecule.

This work is an example of how structural information obtained by using synthetic peptides that recapitulate the binding epitope of a natural partner, can be combined to fragment screening to gauge future optimization of small molecules.

In the last part of the project, I performed initial investigation toward elucidating the binding mode of the PHD-BrD tandem module of two proteins: BAZ1B and TRIM66. These two proteins are both involved in important epigenetic pathways related to diseases, justifying and motivating the interest in studying them.

For the PHD of BAZ1B, conversely to our expectations, our data suggest that the domain alone does not bind to histone tails. The PHD-BrD tandem module did not express in *E.coli* cells suggesting the requirement to express it in a superior system, like insect cells or mammalian cells. Studies on the tandem module would allow investigation on the mutual relationship between the two domains and on the effect of the linker. Indeed, this linker could positively or negatively influence the recognition mode with a potential binding partner.

On the other side, in this work, the tandem module of TRIM66 showed preferences of binding for the unmodified H3 10-mer peptide and H3K9me2, presumably ascribable to the PHD finger. In this tandem, which contains a short linker between the two reader domains, cooperativity of binding was observed for the longer peptide form and in the presence of acetylation on H3K18ac and H3K23ac. Further investigation is warranted in order to explore other meaningful PTMs and to gain insights on the molecular basis of the interactions. Nevertheless, to date, this work provide the first report to our knowledge characterizing in detail the binding between the TRIM66 and the histone tails and establishing the fundamentals blueprints to further proceeding to investigate on this interaction.

In conclusion, the work here presented demonstrated that PHD zinc fingers remain challenging targets to small-molecules but they can also be perceived as a class of ligandable reader domains. This expands the view of what is considered nowadays as a suitable target in the epigenetic field, proposing new reader domains as appealing and affordable target for the drug discovery community. The disclosed structures of these bound fragments together with the information gained about the binding mode of the

PHD fingers with their natural binding partners, have provided important new knowledge that could seed future optimization of these newly identified chemical scaffolds, aiming towards the generation of potent and selective chemical tools targeting PHD zinc finger domains.

CHAPTER 6.

MATERIALS AND METHODS

All the reagents used were purchased by Sigma Aldrich, unless differently specified.

6.1 MOLECULAR BIOLOGY METHODS

6.1.1 Plasmids

Plasmids for the expression of the PHD zinc finger and tandem module PHD zinc finger - Bromodomain (PHD – BrD) of both proteins BAZ2A (Uniprot code: Q9UIF9) and BAZ2B (Uniprot code Q9UIF8) were previously generated in the group by Dr. Cynthia Tallant using the pCri11b plasmid [45]. Each construct was expressed with a 6His tag fused with small ubiquitin like modifier (SUMO)-1 tag at the N terminus of the protein (details of the constructs in Table 6.1).

The cDNAs encoding for the PHD zinc finger and Bromodomain tandem of human BAZ1B (Uniprot code Q9UIG0) was synthetized by Invitrogen (Thermo Fisher Scientific) with codons optimized for *E. coli* expression and supplied into a pMA-T (AmpR) vector. The regions of interest were amplified by PCR using specific primers (Table 1, Appendix) that introduced the NdeI restriction site at 5' terminus and the XhoI restriction site at 3' terminus. Constructs were then subcloned into the pCri11b and pET-28(+) plasmids using the DNA Ligation kit (TaKaRa), following manufacturer's guidelines. The TEV cleavage site was added to those constructs subcloned in pET-28(+).

The synthetic gene of TRIM66 (Uniprot code O15016) was purchased from Invitrogen (Thermo Fisher Scientific). The cDNA, with codons optimized for *E. coli* expression, was supplied into the pMA-T (AmpR) plasmid. Constructs of different length were obtained as described above for BAZ1B constructs and, subsequently, subcloned into pCri11b and pET-28a(+) plasmids using the DNA Ligation kit (TaKaRa). Table 2 in Appendix 1 reports details of the primers used.

In Table 6.1 a summary of all the constructs relevant to the described studies is provided.

Plasmid ID	Protein	MW(Da)	Tags/cleavage site	Vector	Antibiotic
CTA007	PHD BAZ2A(1673-1728)	6589	6His-SUMO/SENP1	pCri11b	Amp
CTA006	TANDEM BAZ2A(1673-1899)	26533	6His-SUMO/SENP1	pCri11b	Amp
CTA016	PHD BAZ2B(1928-1983)	6534	6His-SUMO/SENP1	pCri11b	Amp
CTA014	TANDEM BAZ2B(1928-2168)	28517	6His-SUMO/SENP1	pCri11b	Amp
AAM005	PHD BAZ1B(1182-1237)	6624	6His-SUMO/SENP1	pCri11b	Amp
AAM006	TANDEM BAZ1B(1185-1444)	30439	6His-SUMO/ SENP1	pCri11b	Amp
AAM014	TANDEM BAZ1B(1182-1443)	30640	6His-SUMO/ SENP1	pCri11b	Amp
AAM009	BrD BAZ1B(1321-1449)	15408	6His-SUMO/ SENP1	pCri11b	Amp
AAM010	BrD BAZ1B(1310-1449)	16654	6His-SUMO/ SENP1	pCri11b	Amp
AAM015	PHD BAZ1B(1182-1237)	6442	6His/Thrombin _TEV	pET28-a(+)	Kan
AAM021	BrD BAZ1B(1338-1444)	12780	6His/Thrombin _TEV	pET28-a(+)	Kan
AAM022	BrD BAZ1B(1331-1444)	13593	6His/Thrombin _TEV	pET28-a(+)	Kan
AAM007	BrD BAZ1B(1321-1449)	15227	6His/Thrombin _TEV	pET28-a(+)	Kan
AAM008	BrD BAZ1B(1310-1449)	16474	6His/Thrombin _TEV	pET28-a(+)	Kan
AAM012	PHD TRIM66(970-1018)	5613	6His-SUMO/ SENP1	pCri11b	Amp
AAM013	BrD TRIM66(1041-1150)	13922	6His-SUMO/ SENP1	pCri11b	Amp
AAM011	TANDEM TRIM66(970-1150)	20886	6His-SUMO/ SENP1	pCri11b	Amp

AAM001	TANDEM	TRIM66$_{(970-1150)}$	20886	6His/Thrombin_TEV	pET28-a(+)	Kan
AAM004	TANDEM	TRIM66$_{(966-1150)}$	21340	6His/Thrombin_TEV	pET28-a(+)	Kan
AAM002	TANDEM	TRIM66$_{(966-1159)}$	22481	6His/Thrombin_TEV	pET28-a(+)	Kan
AAM003	TANDEM	TRIM66$_{(966-1168)}$	23417	6His/Thrombin_TEV	pET28-a(+)	Kan

Table 6.1. Summary of the constructs produced in this thesis work.

pET28-a(+) contains double cleavage site as reported in column four: Thrombin and TEV cleavage sites. MW refers to the cleaved domains.

6.1.2 Transformations

Different chemically competent *E.coli* cell strains were used for the transformation process.

DH5-α strain was used for DNA amplification; BL21(DE3), Rosetta, C41(DE3) and ArticExpress(DE3) were used for labelled and unlabelled protein expression. The protocol used for transformation was the same for all the strains. Briefly, 100 µL of competent cells were thawed on ice. To these 1 µL of plasmid DNA (~100 ng) was added and incubated for 10 min on ice. Heat shock was performed for 90 s at 42°C and then cooled on ice for 10 min. Cells were subsequently diluted with 300 µL of Lysogeny broth (LB) and incubated at 37°C for 1 h, then plated onto LB/agar plates, containing the relevant antibiotic, and incubated overnight at 37°C.

6.1.3 Site-directed mutagenesis

Site-directed mutagenesis on PHD of BAZ2A/B were performed by PCR using the Phusion DNA Polymerase (Thermo Fisher Scientific) and specific pairs of primers (Table 3, Appendix) that introduced the desired mutation [46]. The PCR amplification product was incubated with Dpn I (New England BioLabs) for 1 h at 37°C, which recognizes and digest DNA carrying methylated sites on the parental DNA strands. The DNA was subsequently

transformed in DH5-α *E. coli* strains. Single colonies were picked to inoculate 5 mL of LB plus 100 µg/mL ampicillin and grown overnight at 37°C. The following day, DNA was extracted from the bacterial cultures using the QIAprep Spin Miniprep Kit (Qiagen) and the presence of the desired mutation was checked by DNA sequencing.

6.1.4 Protein expression

PHD zinc finger and tandem constructs of BAZ2A/B and PHD zinc finger constructs of BAZ1B were expressed in BL21 (DE3) cells. Colonies from freshly transformed cells were grown in 50 mL LB containing 100 µg/mL ampicillin and incubated overnight at 37°C with 200 rpm shaking. The following morning a dense culture was obtained and this was diluted in order to reach OD_{600} ~ 0.05 in fresh LB medium supplemented with 100 µg/mL ampicillin. The culture was then incubated at 37 °C with 200 rpm shacking until the OD_{600} reached ~ 0.6; then the temperature was decreased to 20 °C and protein overexpression was induced with 200 µM IPTG (Isopropyl β-D-1-thiogalactopyranoside) under 200 rpm shacking for 16 h. Cells were harvested by centrifugation at 4200 rpm for 30 min at 4°C using a Beckman Coulter Avanti J-6 MC centrifuge.
The same protocol was adopted for the expression of the PHD-BrD tandem construct of TRIM66 except for the antibiotic used that was kanamycin at 50 µg/mL and for the induction that was performed at an OD_{600} of ~1 with 300 µM IPTG, decreasing the temperature at 16°C.

6.1.5 Isotopically-labelled protein expression

For the expression of isotopically labelled ^{15}N PHD of BAZ2A/B was used a minimal media M9 [185]. M9 media contained: 6.8 g/L Na2HPO4, 3 g/L KH2PO4, 0.5 g/L NaCl, 0.044 g/L CaCl2, 0.24 g/L MgSO4, 16 mg/L MnCl2, 5 mg/L FeCl3, 0.5 mg/L $ZnCl_2$, 0.1 mg/L $CuCl_2$, 0.1 mg/L $CoCl_2$, 0.1 mg/mL H_3BO_3 and 0.05 g/L EDTA, 0.04 g/L Na_2SO_4, 0.24 g/L $MgSO_4$, 0.044 g/L $CaCl_2$, 0.001 mg/mL Biotin and Thiamine supplemented with 1 g/L $(^{15}NH_4)_2SO_4$ (Goss Scientific) as the sole nitrogen source.

The same minimal media was used for the expression of isotopically labelled $^{15}N/^{13}C$ PHD

of BAZ1B where the sole source of nitrogen and carbon were 1 g/L $(^{15}NH_4)_2SO_4$ (Goss Scientific) and 2 g/L ^{13}C D-glucose (Goss Scientific).

6.1.6 Protein purification

The purification of the 6His-SUMO tagged PHD zinc finger and PHD-BrD tandem of BAZ2A/B was performed as previously described [45]. Following protein expression, cell cultures were harvested by centrifugation and the bacterial pellets was re-suspended in Lysis Buffer (20 mM HEPES, 500 mM NaCl, 30 mM Imidazole, 1 mM DTT at pH 8) supplemented with 5 mM $MgCl_2$, 0.1 mg/mL of DNAse (Roche) and protease inhibitor cocktail tablets (Roche) (1 tablet for 50 mL per lysate). The cellular lysis was conducted with a high-pressure homogenizer (French Press) and the lysate was cleared by centrifugation at 40000 x *g* for 40 min at 4°C using a Beckman Avanti J-25 centrifuge. The 6His-SUMO fusion protein, present in the soluble fraction, was first purified by affinity chromatography by loading the clear lysate on a nickel-charged column (HisTrap FF crude, GE Healthcare). Then the column was previously washed with 50 mL of lysis buffer and the 6His-SUMO protein eluted with lysis buffer containing 500 mM imidazole. The eluted protein was incubated with SENP1 protease (0.5 mg/ml) overnight at 4°C to cleave the 6His-Sumo tag and at the same time dialyzed to reduce the imidazole concentration to ~5 mM. In the second purification step an additional affinity chromatography on a nickel-charged column was performed to separate the 6His-SUMO tag from the desired protein, collecting the latter in the initial flow through and eluting the tag with 500 mM imidazole buffer. Following overnight dialysis in 20 mM Tris, 5 mM DTT pH 8.0 to reduce the buffer salinity of the sample, the pool fraction was loaded on a cation exchange column (Resource S 6 mL, GE Healthcare) and eluted by increasing the ionic strength of the buffer implemented with 500 mM NaCl. All domain protein constructs were further purified by size exclusion chromatography (HiLoad Superdex 75 16/600, GE Healthcare) in 20 mM Tris, 200 mM NaCl and 5 mM DTT at pH 8.0.

The PHD-BrD tandem constructs of TRIM66 were purified following the same procedure described above except for the buffer used for the Ni-affinity chromatography steps, that was 50 mM $H_2PO_4^-/HPO_4^{2-}$ pH 7.5, 500 mM NaCl, 30 mM Imidazole, 7 mM β-mercaptoethanol, and for the usage of TEV protease (0.1 mg/mL) to cleave the tag.

6.1.7 Protein biotinylation

Biotinylation of the PHD domains of BAZ2A/B was performed using the EZ-Link NHS–PEG_4–biotin kit (Thermo Scientific) following the manufacturer guidelines. BAZ2A/B PHD was mixed in an equimolar ratio with biotin and incubated at room temperature for 1 h. The reaction was performed in a free amine buffer (20 mM HEPES, 100 mM NaCl, 5 mM DTT pH 7.5) in order to avoid competing NHS-ester reactions. Unreacted biotin was quenched by adding 2 µL of 1 M Tris buffer at pH 7.5. To remove any unreacted NHS-biotin the sample was run over a PD-10 desalting column (GE Healthcare) into 50 mM HEPES, pH 7.5, 150 mM NaCl and 1 mM DTT.

6.1.8 Protein methylation

The PHD zinc finger domain of BAZ1B was methylated following the protocol reported from the Oxford Protein Production Facility (OPPF) and the Division of Structural Biology at Oxford University [3]. The reaction was a reductive methylation conducted overnight at 4°C in 50 mM Hepes pH 7.5, 250 mM NaCl at the protein concentration of ~ 0.85 mg/mL (130 µM) in presence of dimethylamine-borane complex and formaldehyde. The methylated protein was then purified by size-exclusion chromatography using a Superdex 75 16/600 column pre-equilibrated in 20 mM Tris-HCl pH 7.5 and 200 mM NaCl. The purification step was important for the quenching of the unreacted starting material, the separation of the product from the reagents and to further monitor the oligomeric state of the protein itself.

6.1.9 DNA and protein concentration

DNA and protein concentrations were determined measuring absorbance at the wavelength of 260 nm and 280 nm, respectively. Measurements were performed using a Nanodrop Microliter UV/Vis spectrophotometer (ThermoScientific). DNA purity was estimated by measuring the A260/A280 ratio and all the samples used had A260/A280 > 1.7. Protein concentration was calculated using the theoretical extinction coefficient predicted by ProtParam [181].

6.1.10 Protein analysis

Chromatographic fractions showing absorbance at 280 nm were analysed by SDS-page gels with 16% or 20% acrylamide, and visualized using Comassie staining (InstantBlue protein stain, Expedeon). Protein identity was confirmed by electrospray mass spectrometry analysis (Fig. A1- A5, Appendix).

6.2 NUCLEAR MAGNETIC RESONANCE (NMR) EXPERIMENTS

All NMR experiments were performed using a AV-500 MHz Bruker spectrometer equipped with a 5 mm CPQCI-^{1}H-^{19}F /^{13}C/^{15}N/D Z-GRD cryoprobe.

6.2.1 NMR backbone assignment

PHD zinc finger BAZ1B backbone resonance assignment was performed using 2D ^{15}N(^{1}H)-HSQC and 3D HNCO, HN(CA)CO, CBCA(CO)NH and HNCACB. Spectra were recorded on 500 µL sample in a Shigami 5 mm symmetrical tube. The sample contained ^{15}N/^{13}C PHD of BAZ1B at a concentration of 500 µM in NMR buffer (20 mM $H_2PO_4^-$/HPO_4^{2-} pH 6.3, 50 mM KCl, 10% D_2O and 1 mM DTT).

Acquisition times for the 3D experiments were: 7-8 ms (^{13}C), 12-16 ms (^{15}N) and 100-127 ms (^{1}H).

Acquisition times for the HSQC experiments were: 60 ms (^{15}N) and 120 ms (^{1}H).

All spectra were acquired and processed with TopSpin (Bruker) and analysed with CcpNmr *[154]*.

6.2.2 Chemical shift perturbation (CSP) experiments

(^{15}N-^{1}H)-HSQC spectra were recorded on 200 µL sample, in 3-mm NMR capillary tube, of ^{15}N-labelled BAZ1B PHD at a concentration of 150 µM in NMR buffer (20 mM $H_2PO_4^-$ /HPO_4^{2-} pH 6.3, 50 mM KCl and 1 mM DTT and 20% D_2O) in absence or presence of 20-mer N-terminal histone peptides at increasing concentrations. Spectra were recorded

on samples with the following protein : peptide molar ratio 1:0, 1:1, 1:2, 1:4, 1:8 and 1:10.

^{15}N(^{1}H)-HSQC spectra of both PHD domains of BAZ2A/B were recorded at sample concentration respectively of ~100 µM for BAZ2A and ~150 µM for BAZ2B in 200 µL NMR buffer as above (pH 6.9 and 6.5, respectively). Spectra were obtained for each protein at increasing concentration of fragment compound in a range between 0.5 mM and 5 mM. The reference spectrum was recorded on the apo form.

In chemical CSP experiments the weighted chemical shift difference ($\Delta\delta_{weighted}$) was calculated with the following equation: $\Delta\delta\text{weighted} = \sqrt{|\Delta\delta\text{H}|^2 + |\Delta\delta\text{N}|^2 * 0.15}$ [186], where ΔδH is the change in chemical shift of the proton and ΔδN is the change in chemical shift of the nitrogen, which is scaled with a factor 0.15 to account for the difference in the range of amide proton and amide nitrogen chemical shifts. The chemical shift changes for each backbone amide group were measured from the peak detected in the HSQC spectrum recorded on the apo form to the peak located at the end of the titration.

Affinities of compounds for PHD of BAZ2A/B were derived from HSQC CSPs and the relative K_D were calculated using the equation below [186]:

$$\Delta\delta\,\text{obs} = \Delta\delta\,\text{max}\frac{\left\{Kd+Pt+Lt-\sqrt{(Kd+Pt+Lt)^2-4Pt\cdot Lt}\right\}}{2[P]t}$$

Where [P]t and [L]t are, respectively, the total concentration of protein and ligand; $\Delta\delta_{obs}$ is the change observed in the shift from the reference spectrum, while $\Delta\delta_{max}$ is the maximum shift change obtained upon saturation and is generally calculated from the fitting , since it is not always possible to measure it experimentally.

For each compound the K_D was extrapolated as a median value of 4–7 different resonances ± s.e.m. as spread of values measured.

Acquisition times for the HSQC experiments were: 60 ms (^{15}N) and 120 ms (^{1}H).

All spectra were recorded and processed with TopSpin (Bruker) and analysed with CCPNMR [5].

6.2.3 Ligand-observed NMR

CPMG (Carr-Purcell-Meiboom-Gill) [126] experiments were performed with different spin-lock time before acquisition, as explained in detail in chapter 3.4. WaterLOGSY [125] used a standard protocol implying a 20ms Gaussian 180° pulse to invert water signal and 1.2 sec for NOE mixing time. STD (Saturation Transfer Distance) [124] experiments were performed alternating 'on' and 'off' resonance respectively at 100 ppm and 1 ppm with a 40 ms selective Gaussian 180° pulse. All the experiments were performed at 25°C in a 3-mm capillary tube (Bruker) containing 200 µL of NMR buffer (20 mM HEPES, 50 mM KCl, 1 mM DTT at pH 6. 9 for PHD of BAZ2A and 6.5 for PHD of BAZ2B) supplemented with 20 µM TSP, 20% D2O and 1 mM fragment in presence or absence of 30 µM protein. NMR spectra were processed and analysed using TopSpin software (Bruker).

6.3 PEPTIDE SYNTHESIS

Peptide synthesis was performed using standard automated solid-phase synthesis on a ResPep SL peptide synthesizer (Intavis) in a 24-column set up. The synthetic protocol was set up on several cycles of deprotection, washing, coupling and washing. Each cycles started with Fmoc deprotection of the N–protected amino acid using 20% piperidine in DMF; successively, coupling was performed in two steps with 1 eq of 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) and 1 eq of 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium-3-oxid-hexafluorophosphate (HATU). To cap any unreacted amino acid, a solution of 5% (v/v) acetic anhydride and N-methylmorpholine (NMM) in DCM was used. Washing was performed with DCM. The solid support phase was a Rink Amide AM resin (200-400 mesh). Peptides were cleaved from the resin by incubation of the resin for 3 h with cleavage mixture containing 92.5% trifluoroacetic acid (TFA), 5% Triisopropylsilane and 2.5% water (v/v/v) (1mL of cleavage mixture for 100-150 mg of resin, as suggested by the manufacturer). Peptides were separated from the resin using a single-fritted column and precipitated in 5 ml of ice-cold diethyl ether. The resulting pellet was washed three times with diethyl ether and, subsequently, dissolved in water and lyophilised in a centrifugal evaporator (Genevac EZ-2 series, SP Scientific).

Amino acids and resin used in the synthesis were purchased by Merck Millipore.

6.3.1 Peptide purification and analysis

Peptides were purified using high-performance liquid chromatography (HPLC) on a Gilson Preparative HPLC System with Waters X-Bridge C18 column (100 mm x 19 mm; 5 µm particle size) at 25 mL/min. Different gradients in a range of 5–25% acetonitrile in water with 0.1% TFA were applied for the purification. All the peptides were retained and eluted at the beginning of the gradient. Removal of TFA was performed using the VAriPure IPE column (Agilent). Absence of TFA was confirmed by ^{19}F NMR spectra.
Purified peptides were identified and analysed by LC-MS on an Agilent Technologies 1200 series HPLC connected to an Agilent Technologies 6130 quadrupole LC/MS linked to an Agilent dioden array detector. Chromatographic runs were performed with a Waters X-Bridge C18 column (50 mm × 2.1 mm, 3.5 µm particle size) or with a column Agilent ZORBAX StableBond 80Å C18 (4.6 x 250 mm, 5 µm particle size). The mobile phase used was 5-95% water/acetonitrile +0.1% TFA over 3 min or 20 min.

6.4 BIOPHYSICAL ASSAYS

6.4.1 Bio-layer interferometry (BLI)

BioLayer Interferometry (BLI) experiments were performed in an Octet RED384 instrument (FòrteBio). In the assay, BAZ1B PHD was screened against a library of N-terminal H3 and H4 derivate peptides carrying multiple post-translational modifications and biotinylated at the C-terminus (Altabioscience, Birmingham, UK). Peptides were immobilized on Super Streptavidin (SSA) biosensor at a concentration of 5 µM. The assays were conducted at 25 °C, in a buffer containing 20 mM HEPES, 200 mM NaCl, 2 mM DTT and pH 8.0, using BAZ1B PHD at a concentration of 20 µM and 50 µM. The experiments were performed in black solid 384-well plate under agitation at 1000 rpm. Cycle steps for the analyses included 120 s of biosensor baseline equilibration, 240 s for associations in wells containing BAZ1B PHD and 240 s for the dissociations in wells

containing only buffer. Data were processed and analysed on the Octet Data Analysis Software, provided by the manufacturer.

6.4.2 Surface Plasmon Resonance (SPR)

SPR experiments were performed using a Biacore T100 instrument (GE Healthcare) at 20 °C. All the experiments were set up in SPR buffer (10 mM HEPES, 150 mM NaCl, 2 mM TCEP, 1% DMSO, 0.005% Tween P20 at pH 7.4).

Protein immobilization onto the sensor chip was performed by injecting a volume of 66 µL of protein at 0.02 mg/mL concentration on the surface of a chip covered with carboxymethylated dextran matrix pre-immobilized with nitrilotriacetic acid (NTA sensor chip form GE Healthcare). Prior to protein immobilization, the chip surface was equilibrated with 20 µL of a 500 µM Ni^{2+} solution in order to pre-load the NTA surface with Ni^{2+} metal ion required for the capturing of His-tagged protein. Stabilization of the adhesion of the protein onto the chip was guaranteed by a capture coupling protocol performed using 30 µL of coupling solution [155]. This latter one was made up mixing equal volume of NHS and EDC from the Amine Coupling Kit (GE Healthcare, BR-1000-50).

Sample data were collected for the 6His-SUMO tagged PHD of BAZ2A/B over six fragment concentration points in a 1:2 serial dilution from 900 µM to 28 µM (in SPR buffer). Reference data included the response of the 6His-SUMO tag immobilized alone and these were subtracted from sample data in order to account for unspecific interactions detected by the 6His-SUMO tag alone. Data, whenever possible, were processed using the 1:1 fitting model with the Biacore T100 Evaluation Software (GE Healthcare), provided from the manufacturer.

6.4.3 Differential scanning fluorimetry (DSF)

Differential scanning fluorimetry (DSF) assays were performed using a CFX96 Touch Real-time PCR detection system (Biorad). The assay was performed using a 96-well PCR plate. For the pH buffer screening in each well (40 µL reaction) the final conditions were: 2.5X Sypro Orange (Invitrogen Molecular Probes) and 10 µM protein in different buffer

at different pH in a range between 5 and 9. For the fragment screening performed on the PHD of BAZ2A/B three different concentrations of compound were tested: 3 mM, 5 mM and 15 mM in 5% (v/v) DMSO in a buffer containing 100 mM MES, 50 mM NaCl, pH 6.0.

The assay was conducted by increasing the temperature from 25 °C to 95 °C at a rate of 1 °C per minute. Fluorescence was detected at the end of each interval. The melting temperature (Tm) for each sample was calculated using the excel spreadsheet "DSF analysis" available at ftp://ftp.sgc.ox.ac.uk/pub/biophysics. Fluorescence was plotted at different temperatures and fitted into the Boltzmann equation [132], using the spreadsheet mentioned above, in order to extrapolate the melting points. Each sample was run in triplicate and the Tm values were reported as the mean of three independent measurement ± propagated s.d.

6.4.4 AlphaLisa

AlphaLisa competition assays were set up in a 384-well plates (PerkinElmer, USA) in 100mM HEPES, 50 mM NaCl, 0.1% BSA and 0.02% CHAPS at pH 7.5. In each well were initially dispensed solutions of 10 nM protein, 160 nM H3 double mutant peptide (described above in Chapter 3.7) and fragment at desired concentration in 4% v/v DMSO. Plate was incubated at room temperature for 1 h and then 10 µg/mL (final concentration) of Anti-FLAG AlphaLisa acceptor beads and 10 µg/mL of Streptavidin donor beads were added to the well under low light conditions. The plate was incubated for a further hour at room temperature and subsequently read on a PHERAstar FS plate reader (BMG Labtech) using laser excitation at 680 nm and filter set on emission light at 615 nm. PHD domains of BAZ2A and BAZ2B were tested against eight different fragment concentrations in 1:5 serial dilution and IC_{50} values were extrapolated using the GraphPad Prism 6 (GraphPad Software, USA). Validity of the assay was tested using the Z-factor and S/B ratio.

$$Z' = 1 - \frac{3(\sigma p + \sigma n)}{Ip - In}$$

$$S/_B = \frac{Ip}{In}$$

Where Ip and In are respectively the mean of the highest signal and of the background signal, while σ represent the standard deviation.

6.4.5 Isothermal titration calorimetry (ITC)

Isothermal titration calorimetry (ITC) experiments were performed using ITC200 micro-calorimeter (GE Healthcare) at 298 K, unless otherwise specified, stirring at 750 rpm.
The PHD zinc finger and PHD-BrD tandem of BAZ2A/B used for the ITC experiment were dialyzed overnight against buffer containing 20 mM Tris, 200 mM NaCl, 1 mM TCEP pH 8.0, using the a D-tube dialyzer MWCO 3.5 KDa (Millipore). The PHD-BrD tandem of TRIM66 was dialyzed against a buffer containing 20 mM Tris, 50 mM NaCl, 1 mM TCEP pH 7.5 using D-tube dialyzer MWCO 6-8 KDa (Millipore).

All ITC experiments were carried out in direct mode titrating peptide solutions (0.9 – 1.5 mM) into protein solutions (50-120 µM) loaded in the calorimetric cell, performing one first injection of 0.4 µL (subsequently discarded during data analysis) followed by 19 injections of 2 µL at 120 s time intervals.
For each peptide, a control experiment of titrant into buffer was performed. This data was then subtracted from the related titration of ligand into protein. The resulting data was then fitted to a single-binding-site model to obtain the stoichiometry *n*, the dissociation constant K_D and the enthalpy of binding ΔH using the MicroCal ORIGIN software package. Each titration was repeated in either duplicate or triplicate whenever possible. K_D values reported from replicate titrations were calculated as the mean ± s.e.m. from the values obtained in independent measurements.

6.4.6 Modified Histone array

The Modified histone peptide array is a tool that can be used to screen proteins for interaction with histone tails and their post-translational modification (PTMs). Each array contains up to 384 different peptides resembling the N-terminal tails of the histone proteins H2A, H2B, H3 and H4 with a combination of several PTMs such as acetylation,

methylation, phosphorylation and citrullination. The array was performed following the protocol suggested by the manufacturer (Active Motif).

The arrays were first blocked with a blocking buffer, containing 10 mM Tris-HCl pH 7.4, 0.05% Tween 20, 150 mM NaCl and non-fat dried milk, overnight at 4^{o}C and then incubated with His-tagged protein at room temperature for 2 h (100 nM of His_6-tagged PHD-BrD tandem Trim66 and 400 nM of His_6-tagged PHD of BAZ1B). The array was then washed and incubated for 1 h at room temperature with anti-His antibody and incubated for 1 h with HRP-conjugated secondary antibody. Eventually, the enhanced chemiluminescence (ECL) reaction was performed and the spots visualized on film. All the reagents for the experiments were furnished from the manufacturer. The intensity of the spots and the graphical analysis of the array were carried out using the Array Analyze software (Active Motif).

6.5 PROTEIN CRYSTALLIZATION

6.5.1 Crystallization trials

Several crystallization trials of PHD of BAZ1B and PHD-BrD tandem of Trim66 in free and bound state were conducted at 12 °C, 20 °C and 4 °C by the sitting drop vapour diffusion method using the 96-well crystallization plate (Hampton Research) and the MOSQUITO crystallization robot for the set up. Optimizations of some hits were performed using the 24 well crystallization plate in hanging and sitting drop (Hampton Research). More details about individual procedures and conditions are available in the result chapters 4.1 and 4.2.

6.5.2 Soaking experiments

Crystals of BAZ2A and BAZ2B PHD domains were obtained by mixing equal volume of protein and crystallisation buffer (2.2-2.4 M Na/K phosphate at pH 8.5) at concentration of 6.5–7 mg/ml and 5.5–6 mg/ml was used for BAZ2A and BAZ2B PHD domains, respectively. Crystals were left to grow at 20°C for at least two days [45].

Crystals of both proteins were soaked with 10–20% of different organic solvents (acetone, ethanol, acetonitrile and 1,4-dioxane) at different concentrations of fragments (10–200 mM). The soaking process was conducted for varying times, between 6 to 24 h. Detailed descriptions in chapter 3.9.

Crystals were also soaked with different concentration of H3 10-mer wt and double mutant peptides and H3 3-mer peptide. Peptides were soluble in the crystallization buffer and details of the soaking are in chapter 2.1 and 2.2.

6.5.3 Data collection and structure solving

The X-ray diffraction experiments on the crystals were carried out at the beamlines at the Diamond Light Source (Didcot, UK) and at ESRF synchrotrons. Data sets were acquired at a wavelength of 0.97 Å. Further data set collection was conducted in house using the Rigaku M007HF X-ray generator equipped with Varimax Cu-VHF optics, a Saturn 944HG$^+$ CCD detector and an AFC-11 4-axis partial χ goniometer.
BAZ2A PHD crystallized in a tetragonal system (space group $P4_32_12$) and BAZ2B PHD crystallized in orthorhombic system (space group P222) so for both of them in order to collect a complete data sets with high multiplicity, I ensured to collect images in a rotational range of at least 90°, regardless the spindle axis.

Images were indexed and integrated using XDS [187, 188] for BAZ2A PHD and iMosflm [189] for BAZ2B PHD, then scaled and merged using Aimless [190] from the CCP4 package. R free flags were copied from the respective apo form pdb model (4QF2 for BAZ2A and 4QF3 for BAZ2B) used for the refinement. Structures were solved using isomorphous refinement with Refmac5 [12] and as pdb model was used the pdb of the apo form.

The model was manually built against $2F_o$-F_c and F_o-F_c maps using Coot [191]. F_o-F_c map showed clear electron density to fit the 10-mer peptides, the 3-mer peptide and fragments in the histone pocket. Several rounds of refinement were performed using Refmac5 with TLS groups generated via TLSMD server [192]. Waters were added to the model at the latest stages of the refinement process using Coot [191]. Residues at the C

terminus of the peptide were not visible as well as the very N terminus residues in all the protein chains, consequently they were not modelled.

Modelling of the fragment was performed using the ligand builder in Coot [191]. The topology files of the fragments, for the refinement, were generated using PRODRG2 server [193].

Several cycles of refinement were carried out till R free and R work reached a reasonable values, consistent with the resolution of the dataset, and till further refinements were no longer improving their values.

CHAPTER 7.

REFERENCES

[1] Berger, S. L., Kouzarides, T., Shiekhattar, R., and Shilatifard, A. (2009) An operational definition of epigenetics, *Genes Dev 23*, 781-783.

[2] Bird, A. (2007) Perceptions of epigenetics, *Nature 447*, 396-398.

[3] Turner, B. M. (2009) Epigenetic responses to environmental change and their evolutionary implications, *Philos T R Soc B 364*, 3403-3418.

[4] Bannister, A. J., and Kouzarides, T. (2011) Regulation of chromatin by histone modifications, *Cell Res 21*, 381-395.

[5] Richmond, T. J. (1999) Hot papers - Crystal structure - Crystal structure of the nucleosome core particle at 2.8 angstrom resolution by K. Luger, A.W. Mader, R.K. Richmond, D.F. Sargent, T.J. Richmond - Comments, *Scientist 13*, 15-15.

[6] Rothbart, S. B., and Strahl, B. D. (2014) Interpreting the language of histone and DNA modifications, *Biochimica et biophysica acta 1839*, 627-643.

[7] Dhe-Paganon, S., Syeda, F., and Park, L. (2011) DNA methyl transferase 1: regulatory mechanisms and implications in health and disease, *Int J Biochem Mol Biol 2*, 58-66.

[8] Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D., and Patel, D. J. (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers, *Nat Struct Mol Biol 14*, 1025-1040.

[9] Yun, M. Y., Wu, J., Workman, J. L., and Li, B. (2011) Readers of histone modifications, *Cell Res 21*, 564-578.

[10] Neilson, C., Santos-Rosa, H., Bannister, A., and Kouzarides, T. (2006) Chromatin modifications and their function, *Febs J 273*, 22-22.

[11] Allfrey, V. G., Faulkner, R., and Mirsky, A. E. (1964) Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis, *Proc Natl Acad Sci U S A 51*, 786-794.

[12] Yang, X. J., and Seto, E. (2007) HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention, *Oncogene 26*, 5310-5318.

[13] Eberharter, A., and Becker, P. B. (2002) Histone acetylation: a switch between repressive and permissive chromatin - Second in review series on chromatin dynamics, *Embo Rep 3*, 224-229.

[14] Lan, F., and Shi, Y. (2009) Epigenetic regulation: methylation of histone and non-histone proteins, *Sci China Ser C 52*, 311-322.

[15] Bannister, A. J., Schneider, R., and Kouzarides, T. (2002) Histone methylation: Dynamic or static?, *Cell 109*, 801-806.

[16] Rea, S., Eisenhaber, F., O'Carroll, N., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D., and Jenuwein, T. (2000) Regulation of chromatin structure by site-specific histone H3 methyltransferases, *Nature 406*, 593-599.

[17] Xiao, B., Jing, C., Wilson, J. R., Walker, P. A., Vasisht, N., Kelly, G., Howell, S., Taylor, I. A., Blackburn, G. M., and Gamblin, S. J. (2003) Structure and catalytic mechanism of the human histone methyltransferase SET7/9, *Nature 421*, 652-656.

[18] Zhang, X., Yang, Z., Khan, S. I., Horton, J. R., Tamaru, H., Selker, E. U., and Cheng, X. D. (2003) Structural basis for the product specificity of histone lysine methyltransferases, *Mol Cell 12*, 177-185.

[19] Herrmann, F., Pably, P., Eckerich, C., Bedford, M. T., and Fackelmayer, F. O. (2009) Human protein arginine methyltransferases in vivo - distinct properties of eight canonical members of the PRMT family, *J Cell Sci 122*, 667-677.
[20] Bedford, M. T. (2006) The Family of Protein Arginine Methyltransferases, *Enzymes 24*, 31-50.
[21] Wang, Y., Wysocka, J., Sayegh, J., Lee, Y. H., Perlin, J. R., Leonelli, L., Sonbuchner, L. S., McDonald, C. H., Cook, R. G., Dou, Y., Roeder, R. G., Clarke, S., Stallcup, M. R., Allis, C. D., and Coonrod, S. A. (2004) Human PAD4 regulates histone arginine methylation levels via demethylimination, *Science 306*, 279-283.
[22] Chang, B. S., Chen, Y., Zhao, Y. M., and Bruick, R. K. (2007) JMJD6 is a histone arginine demethylase, *Science 318*, 444-447.
[23] Mosammaparast, N., and Shi, Y. (2010) Reversal of Histone Methylation: Biochemical and Molecular Mechanisms of Histone Demethylases, *Annu Rev Biochem 79*, 155-179.
[24] Bedford, M. T., and Clarke, S. G. (2009) Protein Arginine Methylation in Mammals: Who, What, and Why, *Mol Cell 33*, 1-13.
[25] Trojer, P., and Reinberg, D. (2007) Facultative heterochromatin: is there a distinctive molecular signature?, *Mol Cell 28*, 1-13.
[26] Rossetto, D., Avvakumov, N., and Cote, J. (2012) Histone phosphorylation: a chromatin modification involved in diverse nuclear events, *Epigenetics-Us 7*, 1098-1108.
[27] Sakabe, K., Wang, Z. H., and Hart, G. W. (2010) beta-N-acetylglucosamine (O-GlcNAc) is part of the histone code, *P Natl Acad Sci USA 107*, 19915-19920.
[28] Swamy, M., Pathak, S., Grzes, K. M., Damerow, S., Sinclair, L. V., van Aalten, D. M., and Cantrell, D. A. (2016) Glucose and glutamine fuel protein O-GlcNAcylation to control T cell self-renewal and malignancy, *Nat Immunol 17*, 712-720.
[29] Dehennaut, V., Leprince, D., and Lefebvre, T. (2014) O-GlcNAcylation, an Epigenetic Mark. Focus on the Histone Code, TET Family Proteins, and Polycomb Group Proteins, *Front Endocrinol (Lausanne) 5*, 155.
[30] Muller, S., Filippakopoulos, P., and Knapp, S. (2011) Bromodomains as therapeutic targets, *Expert Rev Mol Med 13*, 1-21.
[31] Zeng, L., and Zhou, M. M. (2002) Bromodomain: an acetyl-lysine binding domain, *Febs Lett 513*, 124-128.
[32] Smith, S. G., and Zhou, M. M. (2016) The Bromodomain: A New Target in Emerging Epigenetic Medicine, *ACS chemical biology 11*, 598-608.
[33] MacDonald, N., Welburn, J. P. I., Noble, M. E. M., Nguyen, A., Yaffe, M. B., Clynes, D., Moggs, J. G., Orphanides, G., Thomson, S., Edmunds, J. W., Clayton, A. L., Endicott, J. A., and Mahadevan, L. C. (2005) Molecular basis for the recognition of phosphorylated and phosphoacetylated histone H3 by 14-3-3, *Mol Cell 20*, 199-211.
[34] Stucki, M., Clapperton, J. A., Mohammad, D., Yaffe, M. B., Smerdon, S. J., and Jackson, S. P. (2005) MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks, *Cell 123*, 1213-1226.
[35] Tamkun, J. W., Deuring, R., Scott, M. P., Kissinger, M., Pattatucci, A. M., Kaufman, T. C., and Kennison, J. A. (1992) Brahma - a Regulator of Drosophila Homeotic Genes Structurally Related to the Yeast Transcriptional Activator Snf2 Sw12, *Cell 68*, 561-572.
[36] Filippakopoulos, P., and Knapp, S. (2014) Targeting bromodomains: epigenetic readers of lysine acetylation, *Nat Rev Drug Discov 13*, 339-358.
[37] Filippakopoulos, P., Picaud, S., Mangos, M., Keates, T., Lambert, J. P., Barsyte-Lovejoy, D., Felletar, I., Volkmer, R., Muller, S., Pawson, T., Gingras, A. C., Arrowsmith, C. H., and Knapp, S. (2012) Histone recognition and large-scale structural analysis of the human bromodomain family, *Cell 149*, 214-231.
[38] Dawson, M. A., Prinjha, R. K., Dittmann, A., Giotopoulos, G., Bantscheff, M., Chan, W. I., Robson, S. C., Chung, C. W., Hopf, C., Savitski, M. M., Huthmacher, C., Gudgin, E., Lugo,

D., Beinke, S., Chapman, T. D., Roberts, E. J., Soden, P. E., Auger, K. R., Mirguet, O., Doehner, K., Delwel, R., Burnett, A. K., Jeffrey, P., Drewes, G., Lee, K., Huntly, B. J. P., and Kouzarides, T. (2011) Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia, *Nature 478*, 529-533.

[39] Nicodeme, E., Jeffrey, K. L., Schaefer, U., Beinke, S., Dewell, S., Chung, C. W., Chandwani, R., Marazzi, I., Wilson, P., Coste, H., White, J., Kirilovsky, J., Rice, C. M., Lora, J. M., Prinjha, R. K., Lee, K., and Tarakhovsky, A. (2010) Suppression of inflammation by a synthetic histone mimic, *Nature 468*, 1119-1123.

[40] Ferguson, F. M., Fedorov, O., Chaikuad, A., Philpott, M., Muniz, J. R. C., Felletar, I., von Delft, F., Heightman, T., Knapp, S., Abell, C., and Ciulli, A. (2013) Targeting Low-Druggability Bromodomains: Fragment Based Screening and Inhibitor Design against the BAZ2B Bromodomain, *J Med Chem 56*, 10183-10187.

[41] Meier, J. C., Tallant, C., Fedorov, O., Witwicka, H., Hwang, S. Y., van Stiphout, R. G., Lambert, J. P., Rogers, C., Yapp, C., Gerstenberger, B. S., Fedele, V., Savitsky, P., Heidenreich, D., Daniels, D. L., Owen, D. R., Fish, P. V., Igoe, N. M., Bayle, E. D., Haendler, B., Oppermann, U. C. T., Buffa, F., Brennan, P. E., Muller, S., Gingras, A. C., Odgren, P. R., Birnbaum, M. J., and Knapp, S. (2017) Selective Targeting of Bromodomains of the Bromodomain-PHD Fingers Family Impairs Osteoclast Differentiation, *ACS chemical biology 12*, 2619-2630.

[42] Schindler, U., Beckmann, H., and Cashmore, A. R. (1993) Hat3.1, a Novel Arabidopsis Homeodomain Protein Containing a Conserved Cysteine-Rich Region, *Plant J 4*, 137-150.

[43] Musselman, C. A., and Kutateladze, T. G. (2009) PHD Fingers Epigenetic Effectors and Potential Drug Targets, *Mol Interv 9*, 314-323.

[44] Liu, L. H., Zhen, X. T., Denton, E., Marsden, B. D., and Schapira, M. (2012) ChromoHub: a data hub for navigators of chromatin-mediated signalling, *Bioinformatics 28*, 2205-2206.

[45] Tallant, C., Valentini, E., Fedorov, O., Overvoorde, L., Ferguson, F. M., Filippakopoulos, P., Svergun, D. I., Knapp, S., and Ciulli, A. (2015) Molecular basis of histone tail recognition by human TIP5 PHD finger and bromodomain of the chromatin remodeling complex NoRC, *Structure 23*, 80-92.

[46] Bortoluzzi, A., Amato, A., Lucas, X., Blank, M., and Ciulli, A. (2017) Structural basis of molecular recognition of helical histone H3 tail by PHD finger domains, *The Biochemical journal 474*, 1633-1651.

[47] Lan, F., Collins, R. E., De Cegli, R., Alpatov, R., Horton, J. R., Shi, X., Gozani, O., Cheng, X., and Shi, Y. (2007) Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression, *Nature 448*, 718-722.

[48] Pena, P. V., Musselman, C. A., Kuo, A. J., Gozani, O., and Kutateladze, T. G. (2009) NMR assignments and histone specificity of the ING2 PHD finger, *Magn Reson Chem 47*, 352-358.

[49] Shi, X. B., Hong, T., Walter, K. L., Ewalt, M., Michishita, E., Hung, T., Carney, D., Pena, P., Lan, F., Kaadige, M. R., Lacoste, N., Cayrou, C., Davrazou, F., Saha, A., Cairns, B. R., Ayer, D. E., Kutateladze, T. G., Shi, Y., Cote, J., Chua, K. F., and Gozani, O. (2006) ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression, *Nature 442*, 96-99.

[50] Li, H. T., Ilin, S., Wang, W. K., Duncan, E. M., Wysocka, J., Allis, C. D., and Patel, D. J. (2006) Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF, *Nature 442*, 91-95.

[51] Xi, Q. R., Wang, Z. X., Zaromytidou, A. I., Zhang, X. H. F., Chow-Tsang, L. F., Liu, J. X., Kim, H., Barlas, A., Manova-Todorova, K., Kaartinen, V., Studer, L., Mark, W., Patel, D. J., and Massague, J. (2011) A Poised Chromatin Platform for TGF-beta Access to Master Regulators, *Cell 147*, 1511-1524.

[52] Shi, X. B., Kachirskaia, I., Walter, K. L., Kuo, J. H. A., Lake, A., Davrazou, F., Chan, S. M., Martin, D. G. E., Fingerman, I. M., Briggs, S. D., Howe, L., Utz, P. J., Kutateladze, T. G., Lugovskoy, A. A., Bedford, M. T., and Gozani, O. (2007) Proteome-wide analysis in Saccharomyces

cerevisiae identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36, *J Biol Chem 282*, 2450-2455.
[53] Xiong, X., Panchenko, T., Yang, S., Zhao, S., Yan, P., Zhang, W., Xie, W., Li, Y., Zhao, Y., Allis, C. D., and Li, H. (2016) Selective recognition of histone crotonylation by double PHD fingers of MOZ and DPF2, *Nat Chem Biol 12*, 1111-1118.
[54] Gallivan, J. P., and Dougherty, D. A. (1999) Cation-pi interactions in structural biology, *P Natl Acad Sci USA 96*, 9459-9464.
[55] Baker, L. A., Allis, C. D., and Wang, G. G. (2008) PHD fingers in human diseases: Disorders arising from misinterpreting epigenetic marks, *Mutat Res-Fund Mol M 647*, 3-12.
[56] Wagner, E. K., Nath, N., Flemming, R., Feltenberger, J. B., and Denu, J. M. (2012) Identification and characterization of small molecule inhibitors of a plant homeodomain finger, *Biochemistry 51*, 8293-8306.
[57] Miller, T. C., Rutherford, T. J., Birchall, K., Chugh, J., Fiedler, M., and Bienz, M. (2014) Competitive binding of a benzimidazole to the histone-binding pocket of the Pygo PHD finger, *ACS chemical biology 9*, 2864-2874.
[58] Dreveny, I., Deeves, S. E., Fulton, J., Yue, B. G., Messmer, M., Bhattacharya, A., Collins, H. M., and Heery, D. M. (2014) The double PHD finger domain of MOZ/MYST3 induces alpha-helical structure of the histone H3 tail to facilitate acetylation and methylation sampling and modification, *Nucleic Acids Res 42*, 822-835.
[59] Oliver, S. S., Musselman, C. A., Srinivasan, R., Svaren, J. P., Kutateladze, T. G., and Denu, J. M. (2012) Multivalent Recognition of Histone Tails by the PHD Fingers of CHD5, *Biochemistry 51*, 6534-6544.
[60] Fischle, W., Wang, Y. M., and Allis, C. D. (2003) Binary switches and modification cassettes in histone biology and beyond, *Nature 425*, 475-479.
[61] Musselman, C. A., and Kutateladze, T. G. (2011) Handpicking epigenetic marks with PHD fingers, *Nucleic Acids Res 39*, 9061-9071.
[62] Tsai, W. W., Wang, Z., Yiu, T. T., Akdemir, K. C., Xia, W., Winter, S., Tsai, C. Y., Shi, X., Schwarzer, D., Plunkett, W., Aronow, B., Gozani, O., Fischle, W., Hung, M. C., Patel, D. J., and Barton, M. C. (2010) TRIM24 links a non-canonical histone signature to breast cancer, *Nature 468*, 927-932.
[63] Ruthenburg, A. J., Li, H. T., Milne, T. A., Dewell, S., McGinty, R. K., Yuen, M., Ueberheide, B., Dou, Y. L., Muir, T. W., Patel, D. J., and Allis, C. D. (2011) Recognition of a Mononucleosomal Histone Modification Pattern by BPTF via Multivalent Interactions, *Cell 145*, 692-706.
[64] Ruthenburg, A. J., Li, H., Patel, D. J., and Allis, C. D. (2007) Multivalent engagement of chromatin modifications by linked binding modules, *Nat Rev Mol Cell Bio 8*, 983-994.
[65] Zhang, X. J., Zhao, D., Xiong, X. Z., He, Z. M., and Li, H. T. (2016) Multifaceted Histone H3 Methylation and Phosphorylation Readout by the Plant Homeodomain Finger of Human Nuclear Antigen Sp100C, *J Biol Chem 291*, 12786-12798.
[66] Baud, M. G. J., Lin-Shiao, E., Cardote, T., Tallant, C., Pschibul, A., Chan, K. H., Zengerle, M., Garcia, J. R., Kwan, T. T., Ferguson, F. M., and Ciulli, A. (2014) Chemical biology. A bump-and-hole approach to engineer controlled selectivity of BET bromodomain chemical probes, *Science 346*, 638-641.
[67] Runcie, A. C., Chan, K. H., Zengerle, M., and Ciulli, A. (2016) Chemical genetics approaches for selective intervention in epigenetics, *Curr Opin Chem Biol 33*, 186-194.
[68] Bamborough, P., Chung, C. W., Demont, E. H., Furze, R. C., Bannister, A. J., Che, K. H., Diallo, H., Douault, C., Grandi, P., Kouzarides, T., Michon, A. M., Mitchell, D. J., Prinjha, R. K., Rau, C., Robson, S., Sheppard, R. J., Upton, R., and Watson, R. J. (2016) A Chemical Probe for the ATAD2 Bromodomain, *Angew Chem Int Edit 55*, 11382-11386.
[69] Sdelci, S., Lardeau, C. H., Tallant, C., Klepsch, F., Klaiber, B., Bennett, J., Rathert, P., Schuster, M., Penz, T., Fedorov, O., Superti-Furga, G., Bock, C., Zuber, J., Huber, K. V., Knapp, S.,

Muller, S., and Kubicek, S. (2016) Mapping the chemical chromatin reactivation landscape identifies BRD4-TAF1 cross-talk, *Nat Chem Biol 12*, 504-510.
[70] Arrowsmith, C. H., Bountra, C., Fish, P. V., Lee, K., and Schapira, M. (2012) Epigenetic protein families: a new frontier for drug discovery, *Nat Rev Drug Discov 11*, 384-400.
[71] Portela, A., and Esteller, M. (2010) Epigenetic modifications and human disease, *Nature biotechnology 28*, 1057-1068.
[72] Milite, C., Feoli, A., Viviano, M., Rescigno, D., Cianciulli, A., Balzano, A. L., Mai, A., Castellano, S., and Sbardella, G. (2016) The emerging role of lysine methyltransferase SETD8 in human diseases, *Clin Epigenetics 8*, 102.
[73] Falkenberg, K. J., and Johnstone, R. W. (2014) Histone deacetylases and their inhibitors in cancer, neurological diseases and immune disorders, *Nat Rev Drug Discov 13*, 673-691.
[74] Jones, P. A., Issa, J. P., and Baylin, S. (2016) Targeting the cancer epigenome for therapy, *Nat Rev Genet 17*, 630-641.
[75] Stresemann, C., and Lyko, F. (2008) Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine, *International journal of cancer 123*, 8-13.
[76] Richon, V. M. (2006) Cancer biology: mechanism of antitumour action of vorinostat (suberoylanilide hydroxamic acid), a novel histone deacetylase inhibitor, *British journal of cancer 95*, S2-S6.
[77] Marks, P. A., and Breslow, R. (2007) Dimethyl sulfoxide to vorinostat: development of this histone deacetylase inhibitor as an anticancer drug, *Nature biotechnology 25*, 84-90.
[78] Finnin, M. S., Donigian, J. R., Cohen, A., Richon, V. M., Rifkind, R. A., Marks, P. A., Breslow, R., and Pavletich, N. P. (1999) Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors, *Nature 401*, 188-193.
[79] Mann, B. S., Johnson, J. R., Cohen, M. H., Justice, R., and Pazdur, R. (2007) FDA approval summary: Vorinostat for treatment of advanced primary cutaneous T-cell lymphoma, *Oncologist 12*, 1247-1252.
[80] Galdeano, C., and Ciulli, A. (2016) Selectivity on-target of bromodomain chemical probes by structure-guided medicinal chemistry and chemical biology, *Future Med Chem 8*, 1655-1680.
[81] Mirguet, O., Gosmini, R., Toum, J., Clement, C. A., Barnathan, M., Brusq, J. M., Mordaunt, J. E., Grimes, R. M., Crowe, M., Pineau, O., Ajakane, M., Daugan, A., Jeffrey, P., Cutler, L., Haynes, A. C., Smithers, N. N., Chung, C. W., Bamborough, P., Uings, I. J., Lewis, A., Witherington, J., Parr, N., Prinjha, R. K., and Nicodeme, E. (2013) Discovery of epigenetic regulator I-BET762: lead optimization to afford a clinical candidate inhibitor of the BET bromodomains, *J Med Chem 56*, 7501-7515.
[82] Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W. B., Fedorov, O., Morse, E. M., Keates, T., Hickman, T. T., Felletar, I., Philpott, M., Munro, S., McKeown, M. R., Wang, Y., Christie, A. L., West, N., Cameron, M. J., Schwartz, B., Heightman, T. D., La Thangue, N., French, C. A., Wiest, O., Kung, A. L., Knapp, S., and Bradner, J. E. (2010) Selective inhibition of BET bromodomains, *Nature 468*, 1067-1073.
[83] Milosevich, N., and Hof, F. (2016) Chemical Inhibitors of Epigenetic Methyllysine Reader Proteins, *Biochemistry 55*, 1570-1583.
[84] Bonetta, L. (2010) Interactome under construction, *Nature 468*, 851-854.
[85] Jones, S., and Thornton, J. M. (1996) Principles of protein-protein interactions, *P Natl Acad Sci USA 93*, 13-20.
[86] Cheng, A. C., Coleman, R. G., Smyth, K. T., Cao, Q., Soulard, P., Caffrey, D. R., Salzberg, A. C., and Huang, E. S. (2007) Structure-based maximal affinity model predicts small-molecule druggability, *Nature biotechnology 25*, 71-75.
[87] Arkin, M. R., Tang, Y. Y., and Wells, J. A. (2014) Small-Molecule Inhibitors of Protein-Protein Interactions: Progressing toward the Reality, *Chem Biol 21*, 1102-1114.

[88] Guo, W. X., Wisniewski, J. A., and Ji, H. T. (2014) Hot spot-based design of small-molecule inhibitors for protein-protein interactions, *Bioorg Med Chem Lett 24*, 2546-2554.
[89] Moreira, I. S., Fernandes, P. A., and Ramos, M. J. (2007) Hot spots-A review of the protein-protein interface determinant amino-acid residues, *Proteins-Structure Function and Bioinformatics 68*, 803-812.
[90] Cardote, T. A. F., and Ciulli, A. (2016) Cyclic and Macrocyclic Peptides as Chemical Tools To Recognise Protein Surfaces and Probe Protein-Protein Interactions, *ChemMedChem 11*, 787-794.
[91] Edfeldt, F. N. B., Folmer, R. H. A., and Breeze, A. L. (2011) Fragment screening to predict druggability (ligandability) and lead discovery success, *Drug Discov Today 16*, 284-287.
[92] Pauling, L. (1946) Molecular architecture and biological reactions, *Chemical & Engineering News 24*, 1375-1377.
[93] Hopkins, A. L., and Groom, C. R. (2002) The druggable genome, *Nat Rev Drug Discov 1*, 727-730.
[94] Sugiyama, Y. (2005) Druggability: selecting optimized drug candidates, *Drug Discov Today 10*, 1577-1579.
[95] Jencks, W. P. (1981) On the Attribution and Additivity of Binding-Energies, *P Natl Acad Sci-Biol 78*, 4046-4050.
[96] Scott, D. E., Coyne, A. G., Hudson, S. A., and Abell, C. (2012) Fragment-Based Approaches in Drug Discovery and Chemical Biology, *Biochemistry 51*, 4990-5003.
[97] Hopkins, A. L., Keseru, G. M., Leeson, P. D., Rees, D. C., and Reynolds, C. H. (2014) The role of ligand efficiency metrics in drug discovery, *Nat Rev Drug Discov 13*, 105-121.
[98] Jones, M. H., Hamana, N., Nezu, J., and Shimane, M. (2000) A novel family of bromodomain genes, *Genomics 63*, 40-45.
[99] Ito, T., Levenstein, M. E., Fyodorov, D. V., Kutach, A. K., Kobayashi, R., and Kadonaga, J. T. (1999) ACF consists of two subunits, Acf1 and ISWI, that function cooperatively in the ATP-dependent catalysis of chromatin assembly, *Gene Dev 13*, 1529-1539.
[100] Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., Cuche, B. A., de Castro, E., Lachaize, C., Langendijk-Genevaux, P. S., and Sigrist, C. J. A. (2008) The 20 years of PROSITE, *Nucleic Acids Res 36*, D245-D249.
[101] Santoro, R., and Grummt, I. (2005) Epigenetic mechanism of rRNA gene silencing: Temporal order of NoRC-mediated histone modification, chromatin remodeling, and DNA methylation, *Mol Cell Biol 25*, 2539-2546.
[102] Santoro, R., Li, J. W., and Grummt, I. (2002) The nucleolar remodeling complex NoRC mediates heterochromatin formation and silencing of ribosomal gene transcription, *Nat Genet 32*, 393-396.
[103] Gu, L., Frommel, S. C., Oakes, C. C., Simon, R., Grupp, K., Gerig, C. Y., Bar, D., Robinson, M. D., Baer, C., Weiss, M., Gu, Z., Schapira, M., Kuner, R., Sultmann, H., Provenzano, M., Cancer, I. P. o. E. O. P., Yaspo, M. L., Brors, B., Korbel, J., Schlomm, T., Sauter, G., Eils, R., Plass, C., and Santoro, R. (2015) BAZ2A (TIP5) is involved in epigenetic alterations in prostate cancer and its overexpression predicts disease recurrence, *Nature genetics 47*, 22-30.
[104] Arking, D. E., Junttila, M. J., Goyette, P., Huertas-Vazquez, A., Eijgelsheim, M., Blom, M. T., Newton-Cheh, C., Reinier, K., Teodorescu, C., Uy-Evanado, A., Carter-Monroe, N., Kaikkonen, K. S., Kortelainen, M. L., Boucher, G., Lagace, C., Moes, A., Zhao, X., Kolodgie, F., Rivadeneira, F., Hofman, A., Witteman, J. C., Uitterlinden, A. G., Marsman, R. F., Pazoki, R., Bardai, A., Koster, R. W., Dehghan, A., Hwang, S. J., Bhatnagar, P., Post, W., Hilton, G., Prineas, R. J., Li, M., Kottgen, A., Ehret, G., Boerwinkle, E., Coresh, J., Kao, W. H., Psaty, B. M., Tomaselli, G. F., Sotoodehnia, N., Siscovick, D. S., Burke, G. L., Marban, E., Spooner, P. M., Cupples, L. A., Jui, J., Gunson, K., Kesaniemi, Y. A., Wilde, A. A., Tardif, J. C., O'Donnell, C. J., Bezzina, C. R., Virmani, R., Stricker, B. H., Tan, H. L., Albert, C. M.,

Chakravarti, A., Rioux, J. D., Huikuri, H. V., and Chugh, S. S. (2011) Identification of a sudden cardiac death susceptibility locus at 2q24.2 through genome-wide association in European ancestry individuals, *Plos Genet 7*, e1002158.

[105] Xiao, A., Li, H., Shechter, D., Ahn, S. H., Fabrizio, L. A., Erdjument-Bromage, H., Ishibe-Murakami, S., Wang, B., Tempst, P., Hofmann, K., Patel, D. J., Elledge, S. J., and Allis, C. D. (2009) WSTF regulates the H2A.X DNA damage response via a novel tyrosine kinase activity, *Nature 457*, 57-62.

[106] Barnett, C., and Krebs, J. E. (2011) WSTF does it all: a multifunctional protein in transcription, repair, and replication, *Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire 89*, 12-23.

[107] Bayes, M., Magano, L. F., Rivera, N., Flores, R., and Jurado, L. A. P. (2003) Mutational mechanisms of Williams-Beuren syndrome deletions, *Am J Hum Genet 73*, 131-151.

[108] Kitagawa, H., Fujiki, R., Yoshimura, K., Oya, H., and Kato, S. (2011) Williams syndrome is an epigenome-regulator disease, *Endocr J 58*, 77-85.

[109] Lalli, M. A., Jang, J. W., Park, J. H. C., Wang, Y. D., Guzman, E., Zhou, H. J., Audouard, M., Bridges, D., Tovar, K. R., Papuc, S. M., Tutulan-Cunita, A. C., Huang, Y. D., Budisteanu, M., Arghir, A., and Kosik, K. S. (2016) Haploinsufficiency of BAZ1B contributes to Williams syndrome through transcriptional dysregulation of neurodevelopmental pathways, *Hum Mol Genet 25*, 1294-1306.

[110] Borden, K. L. B., and Freemont, P. S. (1996) The RING finger domain: A recent example of a sequence-structure family, *Current opinion in structural biology 6*, 395-401.

[111] Short, K. M., and Cox, T. C. (2006) Subclassification of the RBCC/TRIM superfamily reveals a novel motif necessary for microtubule binding, *J Biol Chem 281*, 8970-8980.

[112] Liu, J., Zheng, Q., Deng, Y. Q., Cheng, C. S., Kallenbach, N. R., and Lu, M. (2006) A seven-helix coiled coil, *P Natl Acad Sci USA 103*, 15457-15462.

[113] Reymond, A., Meroni, G., Fantozzi, A., Merla, G., Cairo, S., Luzi, L., Riganelli, D., Zanaria, E., Messali, S., Cainarca, S., Guffanti, A., Minucci, S., Pelicci, P. G., and Ballabio, A. (2001) The tripartite motif family identifies cell compartments, *Embo Journal 20*, 2140-2151.

[114] Khetchoumian, K., Teletin, M., Mark, M., Lerouge, T., Cervino, M., Oulad-Abdelghani, M., Chambon, P., and Losson, R. (2004) TIF1 delta, a novel HP1-interacting member of the transcriptional intermediary factor 1 (TIF1) family expressed by elongating spermatids, *J Biol Chem 279*, 48329-48341.

[115] Liu, J. L., Wu, W. X., Xie, Y. H., Lv, X. D., Ling, D. Y., and Yang, Z. P. (2016) Overexpression of TRIM66 functions as an oncogene in lung cancer progression, *Int J Clin Exp Patho 9*, 4966-4977.

[116] Chen, Y., Guo, Y. F., Yang, H. S., Shi, G. D., Xu, G. H., Shi, J. G., Yin, N., and Chen, D. Y. (2015) TRIM66 overexpresssion contributes to osteosarcoma carcinogenesis and indicates poor survival outcome, *Oncotarget 6*, 23708-23719.

[117] Bortoluzzi, A., and Ciulli, A. (2017) Protein-Based NMR Methods Applied to Drug Discovery, In *Applied Biophysics for Drug Discovery*, pp 153-173, John Wiley & Sons, Ltd.

[118] Williamson, M. P. (2013) Using chemical shift perturbation to characterise ligand binding, *Prog Nucl Mag Res Sp 73*, 1-16.

[119] Ziarek, J. J., Peterson, F. C., Lytle, B. L., and Volkman, B. F. (2011) Binding Site Identification and Structure Determination of Protein-Ligand Complexes by Nmr: A Semiautomated Approach, *Fragment-Based Drug Design: Tools, Practical Approaches, and Examples 493*, 241-275.

[120] Xu, Y., and Matthews, S. (2013) TROSY NMR spectroscopy of large soluble proteins, *Top Curr Chem 335*, 97-119.

[121] Śledź, P., Abell, C., and Ciulli, A. (2012) Ligand-Observed NMR in Fragment-Based Approaches, In *NMR of Biomolecules*, pp 264-280, Wiley-VCH Verlag GmbH & Co. KGaA.

[122] Dias, D. M., Van Molle, I., Baud, M. G., Galdeano, C., Geraldes, C. F., and Ciulli, A. (2014) Is NMR Fragment Screening Fine-Tuned to Assess Druggability of Protein-Protein Interactions?, *ACS medicinal chemistry letters 5*, 23-28.

[123] Ciulli, A. (2013) Biophysical screening for the discovery of small-molecule ligands, *Methods in molecular biology 1008*, 357-388.

[124] Martin-Pastor, M., Vega-Vazquez, M., De Capua, A., Canales, A., Andre, S., Gabius, H. J., and Jimenez-Barbero, J. (2006) Enhanced signal dispersion in saturation transfer difference experiments by conversion to a 1D-STD-homodecoupled spectrum, *Journal of biomolecular NMR 36*, 103-109.

[125] Dalvit, C., Fogliatto, G., Stewart, A., Veronesi, M., and Stockman, B. (2001) WaterLOGSY as a method for primary NMR screening: practical aspects and range of applicability, *Journal of biomolecular NMR 21*, 349-359.

[126] Carr, H. Y., and Purcell, E. M. (1954) Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments, *Phys Rev 94*, 630-638.

[127] Carvalho, A. L., Trincao, J., and Romao, M. J. (2010) X-Ray Crystallography in Drug Discovery, *Ligand-Macromolecular Interactions in Drug Discovery: Methods and Protocols 572*, 31-56.

[128] Blundell, T. L., and Patel, S. (2004) High-throughput X-ray crystallography for drug discovery, *Curr Opin Pharmacol 4*, 490-496.

[129] Hassell, A. M., An, G., Bledsoe, R. K., Bynum, J. M., Carter, H. L., 3rd, Deng, S. J., Gampe, R. T., Grisard, T. E., Madauss, K. P., Nolte, R. T., Rocque, W. J., Wang, L., Weaver, K. L., Williams, S. P., Wisely, G. B., Xu, R., and Shewchuk, L. M. (2007) Crystallization of protein-ligand complexes, *Acta crystallographica. Section D, Biological crystallography 63*, 72-79.

[130] Rich, R. L., and Myszka, D. G. (2007) Higher-throughput, label-free, real-time molecular interaction analysis, *Analytical biochemistry 361*, 1-6.

[131] Ciulli, A. (2013) Biophysical screening for the discovery of small-molecule ligands, *Methods in molecular biology 1008*, 357-388.

[132] Niesen, F. H., Berglund, H., and Vedadi, M. (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability, *Nat Protoc 2*, 2212-2221.

[133] Dai, R., Geders, T. W., Liu, F., Park, S. W., Schnappinger, D., Aldrich, C. C., and Finzel, B. C. (2015) Fragment-Based Exploration of Binding Site Flexibility in Mycobacterium tuberculosis BioA, *J Med Chem 58*, 5208-5217.

[134] Morreale, F. E., Bortoluzzi, A., Chaugule, V. K., Arkinson, C., Walden, H., and Ciulli, A. (2017) Allosteric Targeting of the Fanconi Anemia Ubiquitin-Conjugating Enzyme Ube2T by Fragment Screening, *J Med Chem 60*, 4093-4098.

[135] Bielefeld-Sevigny, M. (2009) AlphaLISA immunoassay platform- the "no-wash" high-throughput alternative to ELISA, *Assay and drug development technologies 7*, 90-92.

[136] Ferguson, F. M., Dias, D. M., Rodrigues, J. P. G. L. M., Wienk, H., Boelens, R., Bonvin, A. M. J. J., Abell, C., and Ciulli, A. (2014) Binding Hotspots of BAZ2B Bromodomain: Histone Interaction Revealed by Solution NMR Driven Docking, *Biochemistry 53*, 6706-6716.

[137] Chen, P., Chaikuad, A., Bamborough, P., Bantscheff, M., Bountra, C., Chung, C. W., Fedorov, O., Grandi, P., Jung, D., Lesniak, R., Lindon, M., Muller, S., Philpott, M., Prinjha, R., Rogers, C., Selenski, C., Tallant, C., Werner, T., Willson, T. M., Knapp, S., and Drewry, D. H. (2016) Discovery and Characterization of GSK2801, a Selective Chemical Probe for the Bromodomains BAZ2A and BAZ2B, *J Med Chem 59*, 1410-1424.

[138] Drouin, L., McGrath, S., Vidler, L. R., Chaikuad, A., Monteiro, O., Tallant, C., Philpott, M., Rogers, C., Fedorov, O., Liu, M., Akhtar, W., Hayes, A., Raynaud, F., Muller, S., Knapp, S., and Hoelder, S. (2015) Structure enabled design of BAZ2-ICR, a chemical probe targeting the bromodomains of BAZ2A and BAZ2B, *J Med Chem 58*, 2553-2559.

[139] Spiliotopoulos, D., Wamhoff, E. C., Lolli, G., Rademacher, C., and Caflisch, A. (2017) Discovery of BAZ2A bromodomain ligands, *Eur J Med Chem 139*, 564-572.
[140] Zhu, J. Y., Ye, Q., Zhang, L. J., Song, Y. N., Zhang, M., Wang, W. H., and Zhang, H. (2017) TRIM66 knockdown inhibits cell growth, but induces cell cycle arrest and apoptosis of hepatocellular carcinoma cells, *Int J Clin Exp Patho 10*, 1030-1040.
[141] McAllister, T. E., England, K. S., Hopkinson, R. J., Brennan, P. E., Kawamura, A., and Schofield, C. J. (2016) Recent Progress in Histone Demethylase Inhibitors, *J Med Chem 59*, 1308-1329.
[142] Jung, M., Philpott, M., Muller, S., Schulze, J., Badock, V., Eberspacher, U., Moosmayer, D., Bader, B., Schmees, N., Fernandez-Montalvan, A., and Haendler, B. (2014) Affinity Map of Bromodomain Protein 4 (BRD4) Interactions with the Histone H4 Tail and the Small Molecule Inhibitor JQ1, *J Biol Chem 289*, 9304-9319.
[143] Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallogr D 66*, 12-21.
[144] Pace, C. N., and Scholtz, J. M. (1998) A helix propensity scale based on experimental studies of peptides and proteins, *Biophysical journal 75*, 422-427.
[145] Klein, B. J., Simithy, J., Wang, X., Ahn, J., Andrews, F. H., Zhang, Y., Cote, J., Shi, X., Garcia, B. A., and Kutateladze, T. G. (2017) Recognition of Histone H3K14 Acylation by MORF, *Structure 25*, 650-654 e652.
[146] Wang, X. Y., Moore, S. C., Laszckzak, M., and Ausio, J. (2000) Acetylation increases the alpha-helical content of the histone tails of the nucleosome, *J Biol Chem 275*, 35013-35020.
[147] Arita, K., Isogai, S., Oda, T., Unoki, M., Sugita, K., Sekiyama, N., Kuwata, K., Hamamoto, R., Tochio, H., Sato, M., Ariyoshi, M., and Shirakawa, M. (2012) Recognition of modification status on a histone H3 tail by linked histone reader modules of the epigenetic regulator UHRF1, *P Natl Acad Sci USA 109*, 12950-12955.
[148] Kostrhon, S., Kontaxis, G., Kaufmann, T., Schirghuber, E., Kubicek, S., Konrat, R., and Slade, D. (2017) A histone-mimicking interdomain linker in a multidomain protein modulates multivalent histone binding, *The Journal of biological chemistry 292*, 17643-17657.
[149] Bird, G. H., Crannell, W. C., and Walensky, L. D. (2011) Chemical synthesis of hydrocarbon-stapled peptides for protein interaction research and therapeutic targeting, *Current protocols in chemical biology 3*, 99-117.
[150] Chang, Y. S., Graves, B., Guerlavais, V., Tovar, C., Packman, K., To, K. H., Olson, K. A., Kesavan, K., Gangurde, P., Mukherjee, A., Baker, T., Darlak, K., Elkin, C., Filipovic, Z., Qureshi, F. Z., Cai, H., Berry, P., Feyfant, E., Shi, X. E., Horstick, J., Annis, D. A., Manning, A. M., Fotouhi, N., Nash, H., Vassilev, L. T., and Sawyer, T. K. (2013) Stapled alpha-helical peptide drug development: a potent dual inhibitor of MDM2 and MDMX for p53-dependent cancer therapy, *Proc Natl Acad Sci U S A 110*, E3445-E3454.
[151] Yin, H., Lee, G. I., Sedey, K. A., Kutzki, O., Park, H. S., Orner, B. P., Ernst, J. T., Wang, H. G., Sebti, S. M., and Hamilton, A. D. (2005) Terphenyl-based bak BH3 alpha-helical proteomimetics as low-molecular-weight antagonists of Bcl-X-L, *J Am Chem Soc 127*, 10191-10196.
[152] Ziarek, J. J., Peterson, F. C., Lytle, B. L., and Volkman, B. F. (2011) Binding site identification and structure determination of protein-ligand complexes by NMR a semiautomated approach, *Methods in enzymology 493*, 241-275.
[153] Hermann, J. C., Chen, Y. S., Wartchow, C., Menke, J., Gao, L., Gleason, S. K., Haynes, N. E., Scott, N., Petersen, A., Gabriel, S., Vu, B., George, K. M., Narayanan, A., Li, S. H., Qian, H., Beatini, N., Niu, L. H., and Gan, Q. F. (2013) Metal Impurities Cause False Positives in High-Throughput Screening Campaigns, *ACS medicinal chemistry letters 4*, 197-200.

[154] Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline, *Proteins 59*, 687-696.

[155] Kimple, A. J., Muller, R. E., Siderovski, D. P., and Willard, F. S. (2010) A capture coupling method for the covalent immobilization of hexahistidine tagged proteins for surface plasmon resonance, *Methods in molecular biology 627*, 91-100.

[156] Etemadzadeh, M. H., Arashkia, A., Roohvand, F., Norouzian, D., and Azadmanesh, K. (2015) Isolation, cloning, and expression of E. coli BirA gene for biotinylation applications, *Advanced biomedical research 4*, 149.

[157] Fairhead, M., and Howarth, M. (2015) Site-specific biotinylation of purified proteins using BirA, *Methods in molecular biology 1266*, 171-184.

[158] Zhang, J. H., Chung, T. D. Y., and Oldenburg, K. R. (1999) A simple statistical parameter for use in evaluation and validation of high throughput screening assays, *J Biomol Screen 4*, 67-73.

[159] Liebschner, D., Afonine, P. V., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C., and Adams, P. D. (2017) Polder maps: improving OMIT maps by excluding bulk solvent, *Acta crystallographica. Section D, Structural biology 73*, 148-157.

[160] Collins, B., Stevens, R. C., and Page, R. (2005) Crystallization Optimum Solubility Screening: using crystallization results to identify the optimal buffer for protein crystal formation, *Acta Crystallogr F 61*, 1035-1038.

[161] Dong, A. P., Xu, X. H., and Edward, A. M. (2007) In situ proteolysis for protein crystallization and structure determination, *Nat Methods 4*, 1019-1021.

[162] D'Arcy, A., Bergfors, T., Cowan-Jacob, S. W., and Marsh, M. (2014) Microseed matrix screening for optimization in protein crystallization: what have we learned?, *Acta crystallographica. Section F, Structural biology communications 70*, 1117-1126.

[163] D'Arcy, A., Villard, F., and Marsh, M. (2007) An automated microseed matrix-screening method for protein crystallization, *Acta crystallographica. Section D, Biological crystallography 63*, 550-554.

[164] Stewart, P. D. S., Kolek, S. A., Briggs, R. A., Chayen, N. E., and Baldock, P. F. M. (2011) Random Microseeding: A Theoretical and Practical Exploration of Seed Stability and Seeding Techniques for Successful Protein Crystallization, *Cryst Growth Des 11*, 3432-3441.

[165] Obmolova, G., Malia, T. J., Teplyakov, A., Sweet, R. W., and Gilliland, G. L. (2014) Protein crystallization with microseed matrix screening: application to human germline antibody Fabs, *Acta crystallographica. Section F, Structural biology communications 70*, 1107-1115.

[166] Ireton, G. C., and Stoddard, B. L. (2004) Microseed matrix screening to improve crystals of yeast cytosine deaminase, *Acta crystallographica. Section D, Biological crystallography 60*, 601-605.

[167] Abuhammad, A., Lowe, E. D., McDonough, M. A., Shaw Stewart, P. D., Kolek, S. A., Sim, E., and Garman, E. F. (2013) Structure of arylamine N-acetyltransferase from Mycobacterium tuberculosis determined by cross-seeding with the homologous protein from M. marinum: triumph over adversity, *Acta crystallographica. Section D, Biological crystallography 69*, 1433-1446.

[168] Walter, T. S., Meier, C., Assenberg, R., Au, K. F., Ren, J. S., Verma, A., Nettleship, J. E., Owens, R. J., Stuart, D. I., and Grimes, J. M. (2006) Lysine methylation as a routine rescue strategy for protein crystallization, *Structure 14*, 1617-1622.

[169] Derewenda, Z. S., and Vekilov, P. G. (2006) Entropy and surface engineering in protein crystallization, *Acta Crystallogr D 62*, 116-124.

[170] Clamp, M., Cuff, J., Searle, S. M., and Barton, G. J. (2004) The Jalview Java alignment editor, *Bioinformatics 20*, 426-427.

[171] Sultana, A., and Lee, J. E. (2015) Measuring protein-protein and protein-nucleic Acid interactions by biolayer interferometry, *Current protocols in protein science 79*, 19 25 11-26.
[172] Pascual, J., Martinez-Yamout, M., Dyson, H. J., and Wright, P. E. (2000) Structure of the PHD zinc finger from human Williams-Beuren syndrome transcription factor, *J Mol Biol 304*, 723-729.
[173] Oppikofer, M., Sagolla, M., Haley, B., Zhang, H. M., Kummerfeld, S. K., Sudhamsu, J., Flynn, E. M., Bai, T., Zhang, J., Ciferri, C., and Cochran, A. G. (2017) Non-canonical reader modules of BAZ1A promote recovery from DNA damage, *Nat Commun 8*, 862.
[174] Plotnikov, A. N., Yang, S., Zhou, T. J., Rusinova, E., Frasca, A., and Zhou, M. M. (2014) Structural insights into acetylated-histone H4 recognition by the bromodomain-PHD finger module of human transcriptional coactivator CBP, *Structure 22*, 353-360.
[175] Zeng, L., Yap, K. L., Ivanov, A. V., Wang, X. Q., Mujtaba, S., Plotnikova, O., Rauscher, F. J., and Zhou, M. M. (2008) Structural insights into human KAP1 PHD finger-bromodomain and its role in gene silencing, *Nat Struct Mol Biol 15*, 626-633.
[176] Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis, *Nat Protoc 10*, 845-858.
[177] Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content, *Bioinformatics 21*, 3433-3434.
[178] Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005) The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins, *J Mol Biol 347*, 827-839.
[179] Vagenende, V., Yap, M. G., and Trout, B. L. (2009) Mechanisms of protein stabilization and prevention of protein aggregation by glycerol, *Biochemistry 48*, 11084-11096.
[180] Bennion, B. J., and Daggett, V. (2003) The molecular basis for the chemical denaturation of proteins by urea, *Proc Natl Acad Sci U S A 100*, 5142-5147.
[181] Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., and Hochstrasser, D. F. (1999) Protein identification and analysis tools in the ExPASy server, *Methods in molecular biology 112*, 531-552.
[182] Dumon-Seignovert, L., Cariot, G., and Vuillard, L. (2004) The toxicity of recombinant proteins in Escherichia coli: A comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3), *Protein expression and purification 37*, 203-206.
[183] Kim, D. E., Chivian, D., and Baker, D. (2004) Protein structure prediction and analysis using the Robetta server, *Nucleic Acids Res 32*, W526-W531.
[184] Barbieri, I., Cannizzaro, E., and Dawson, M. A. (2013) Bromodomains as therapeutic targets in cancer, *Brief Funct Genomics 12*, 219-230.
[185] Marley, J., Lu, M., and Bracken, C. (2001) A method for efficient isotopic labeling of recombinant proteins, *Journal of biomolecular NMR 20*, 71-75.
[186] Williamson, M. P. (2014) Using chemical shift perturbation to characterise ligand binding (vol 73, pg 1, 2013), *Prog Nucl Mag Res Sp 80*, 64-64.
[187] Kabsch, W. (2010) Xds, *Acta Crystallogr D 66*, 125-132.
[188] Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement, *Acta Crystallogr D 66*, 133-144.
[189] Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM, *Acta crystallographica. Section D, Biological crystallography 67*, 271-281.
[190] Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution?, *Acta Crystallogr D 69*, 1204-1214.
[191] Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallogr D 66*, 486-501.

[192] Painter, J., and Merritt, E. A. (2006) TLSMD web server for the generation of multi-group TLS models, *J Appl Crystallogr 39*, 109-111.
[193] Schuttelkopf, A. W., and van Aalten, D. M. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes, *Acta crystallographica. Section D, Biological crystallography 60*, 1355-1363.

APPENDIX

Primer Name	DNA sequence from 5’ to 3’
FW_BAZ1B_PHD(1182)	CTGATCATATGGCGGAAATGAGCGCAGAAAATGCACGT
RV_BAZ1B_ PHD(1237)	CTAGCTCGAGTCACTATTAACGTGCGGTTGCAGGCTG
FW-BrD1B-105-(1338)	CTACCATATGGAGAACCTGTACTTCCAATCCAGCAGCCGTCGTCAG AG
FW-TEV-PHD1B-(1237)	ATCTCATATGGAGAACCTGTACTTCCAATCCGCACGTTGTAAAGTGT GTC
FW_BrD1B_(1310)	CTGATCATATGCATAGCACCCGTCGTAG
FW_BrD1B_(1321)	CTGATCATATGCCTGTTGATGATGCCGAA
RV_BrD1B_(1449)	CTACCTCGAGTCACTAATACGGATGACCAGGCAG
FW_BrD1B_TEV(1310)	CTGATCATATGGAGAACCTGTACTTCCAATCCCATAGCACCCGTCGT AG
FW_BrD1B_TEV(1321)	CTGATCATATGGAGAACCTGTACTTCCAATCCCCTGTTGATGATGCC GAA

Table 1. Primers designed for BAZ1B constructs

Primer Name	DNA sequence from 5’ to 3’
FW-TEV-PHD-T66(970)	CTGATCATATGGAGAACCTGTACTTCCAATCCGAAGATTTTTGTGCCGTTTG
RV-PHD-T66(1150)	CTACCTCGAGTCACTAGGTCAGGCTACGACACA
RV-BRD-T66(1150)	CTACCTCGAGTCACTATTCCGGATAGATCTCTTTCA
FW-TEV-BRD-TRIM66	CTGATCATATGGAGAACCTGTACTTCCAATCCCCTGGTATGCGTGCAAGT
FW-TEV-PHD-TRIM66(1168)	CTGATCATATGGAGAACCTGTACTTCCAATCCCCGATTGAAAATGAAGATTTTT
RV-BRD-TRIM66/189	CTACCTCGAGTCACTATGCAAAGCGTTTTTCCGGA
RV-BRD-T66/194	CTACCTCGAGTCACTATTCTTGACGAGGCTGTGC
RV-BRD-T66/203	CTACCTCGAGTCACTAGCTGCTAACTTCTTCGCTA
FW-BRD-TRIM66	CTGATCATATGAATCCCCTGGTATGCGTGCAAGT
FW-PHD-TRIM66/LONG	CTGATCATATGAATCCCCGATTGAAAATGAAGATTTTT

Table 2. Primers designed for TRIM66 constructs

Primer Name	DNA sequence from 5’ to 3’
FW-BAZ2A-PHD-D1688N/E1689Q	CGCAAAGGCGATAATAATCAGTTTCTGCTGCTGTGTGAT
RV-BAZ2A-PHD-D1688N/E1689Q	ATCACACAGCAGCAGAAACTGATTATTATCGCCTTTGCG

Table 3. Primers designed for mutation on PHD of BAZ2A.

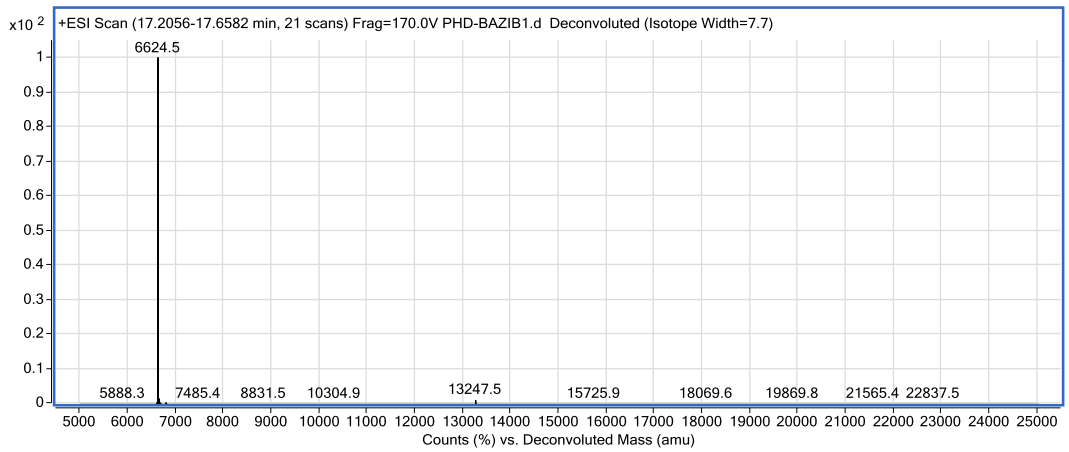


Figure A1.
Deconvoluted electrospray mass spectrum of PHD of BAZ1B.

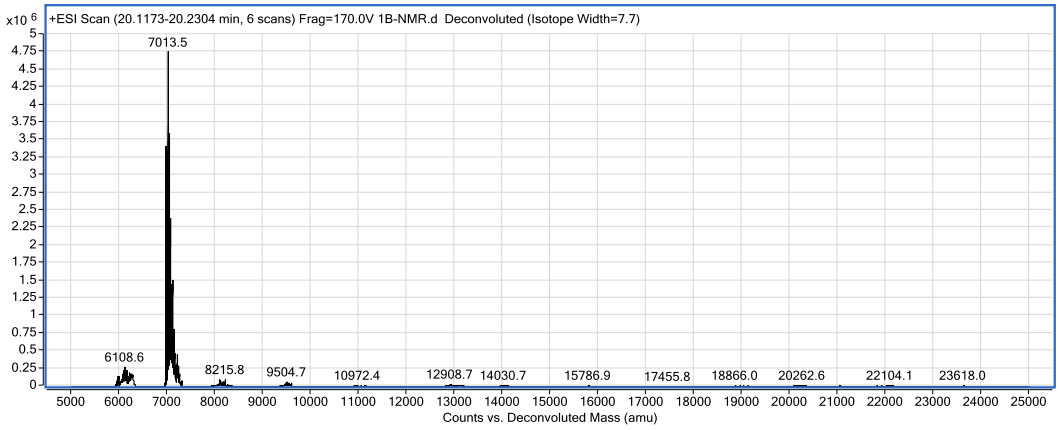


Figure A2.
Deconvoluted electrospray mass spectrum of $^{15}N/^{13}C$ PHD of BAZ1B.

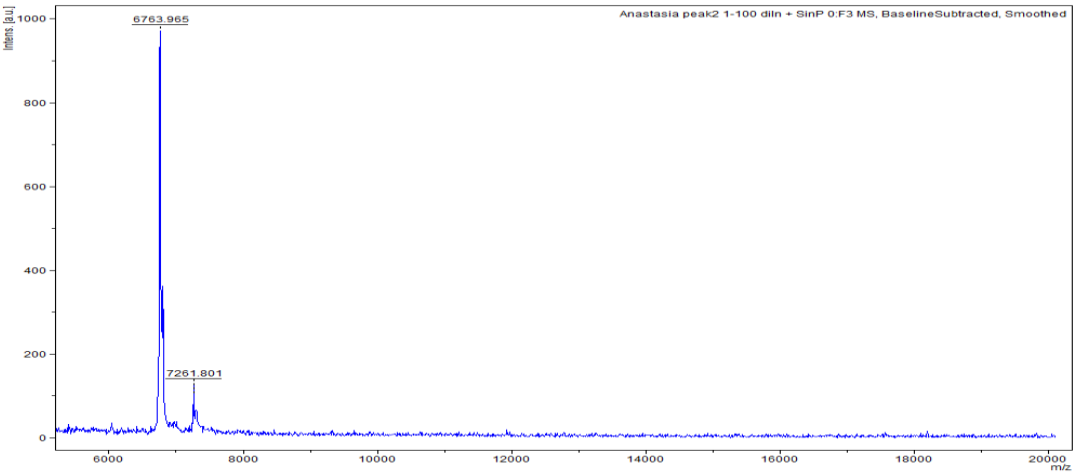


Figure A3.
Deconvoluted electrospray mass spectrum of PHD of BAZ1B methylated at Lys residues.

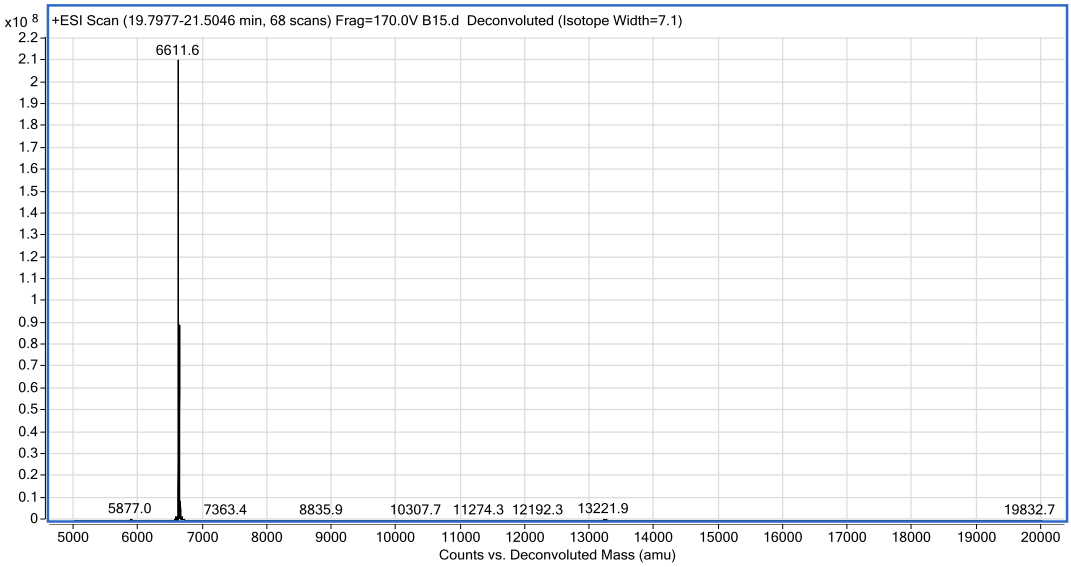


Figure A4.
Deconvoluted electrospray mass spectrum of ^{15}N PHD of BAZ2B.

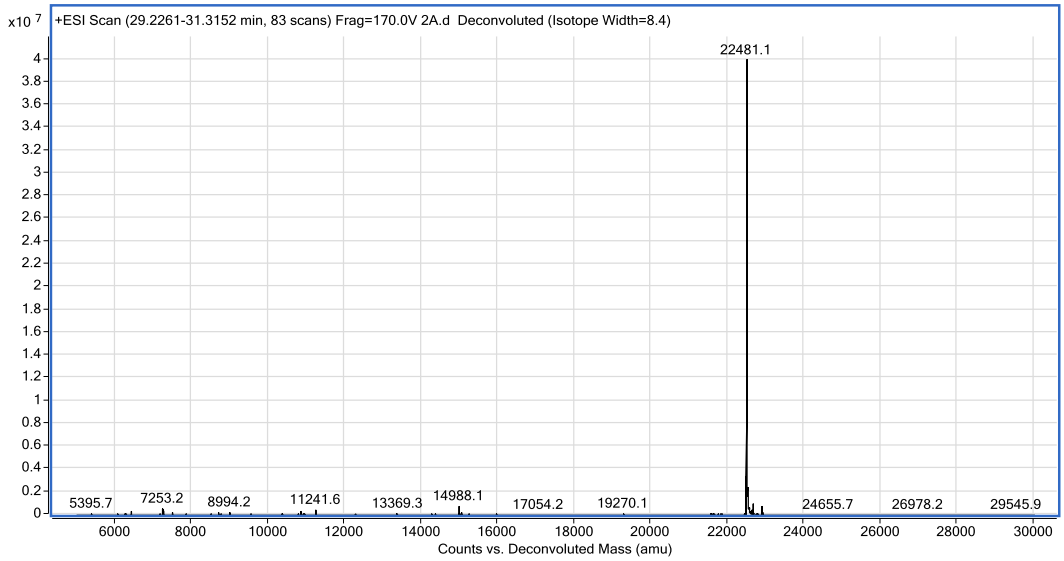


Figure A5.
Deconvoluted electrospray mass spectrum of Trim66/194.

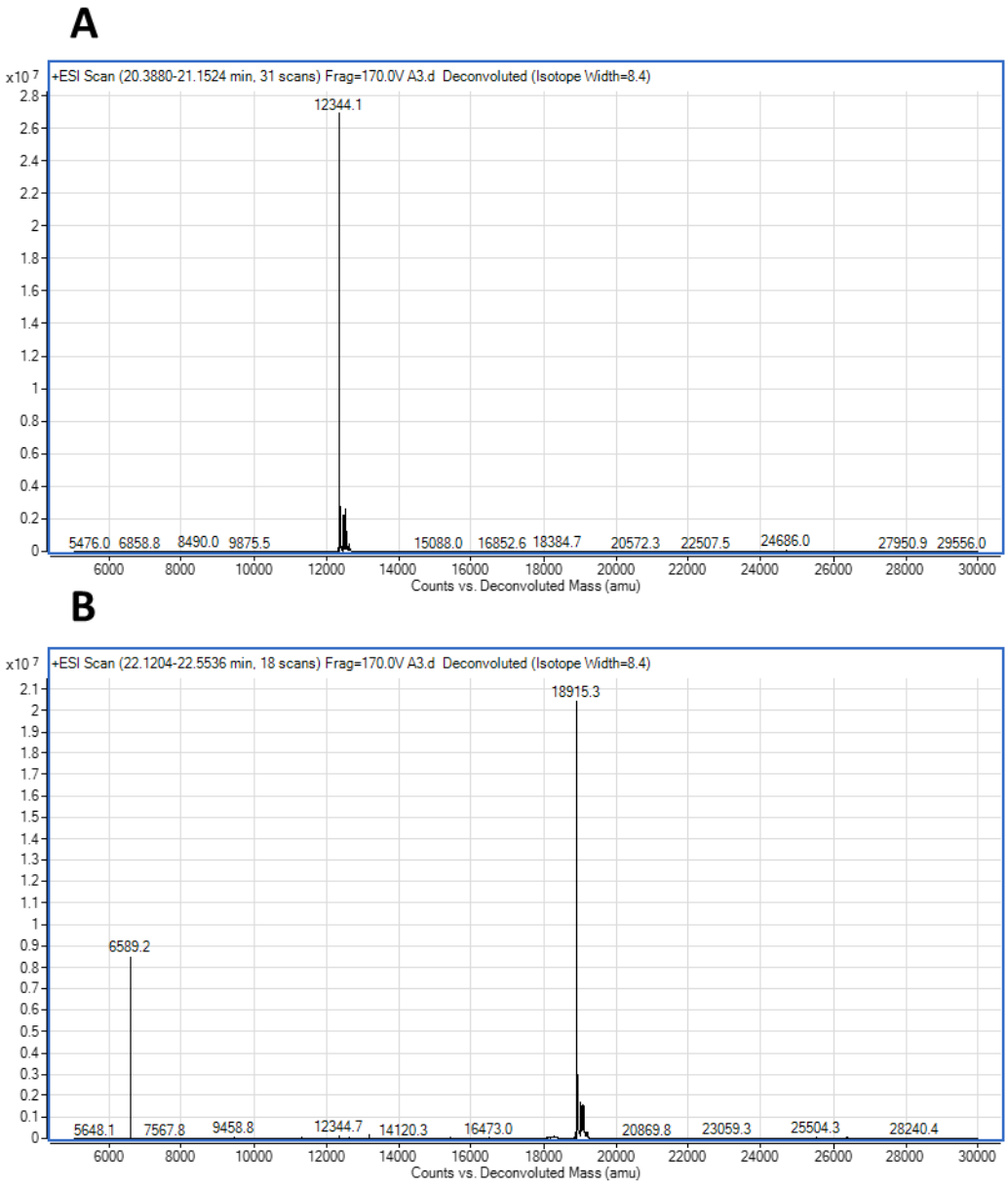


Figure A6. MS analysis of BAZ2A 6His-Sumo-PHD

Deconvoluted electrospray mass spectrum of 6His-Sumo-PHD of BAZ2A. A) Mass relative to MW of 6His-SUMO. B) Mass relative to MW of PHD of BAZ2A and 6His-Sumo-PHD.

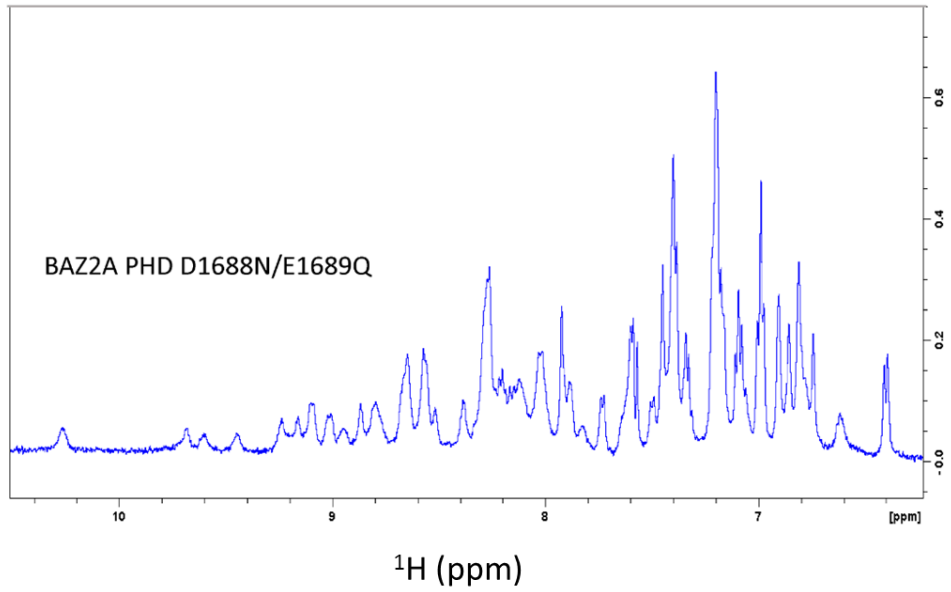


Figure A.7. 1D NMR of BAZ2A PHD D1688N/E1689Q

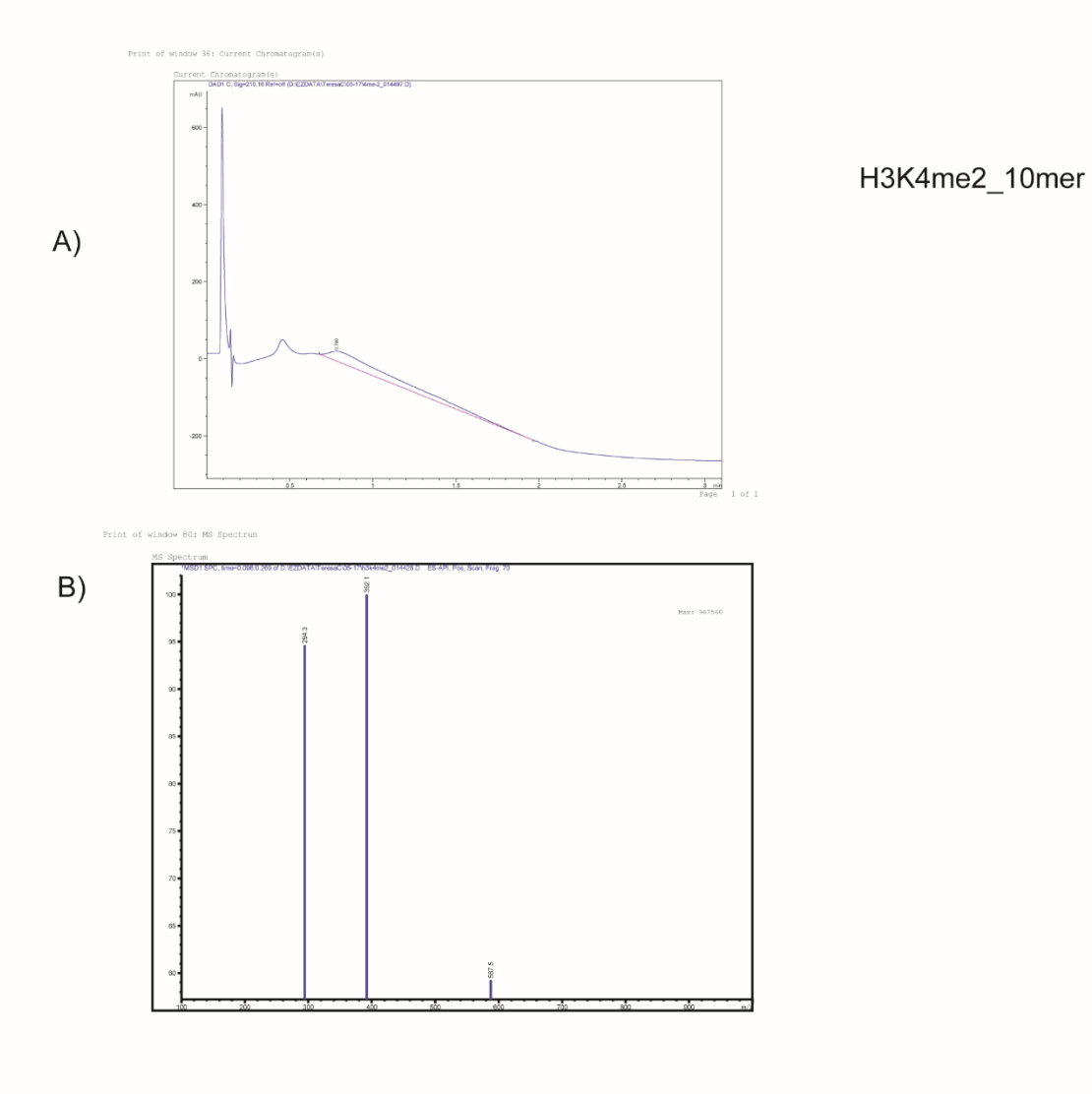
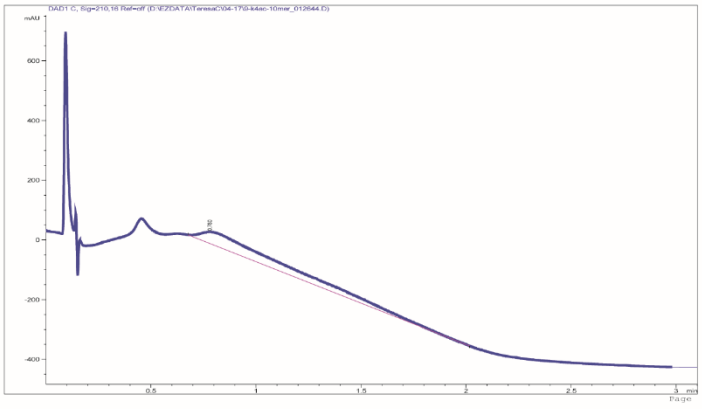


Figure A8.1

LC-MS H3K4me2 10-mer (MW ~1226 Da)

A)



H3K4ac_10mer

Print of window 80: MS Spectrum

B)

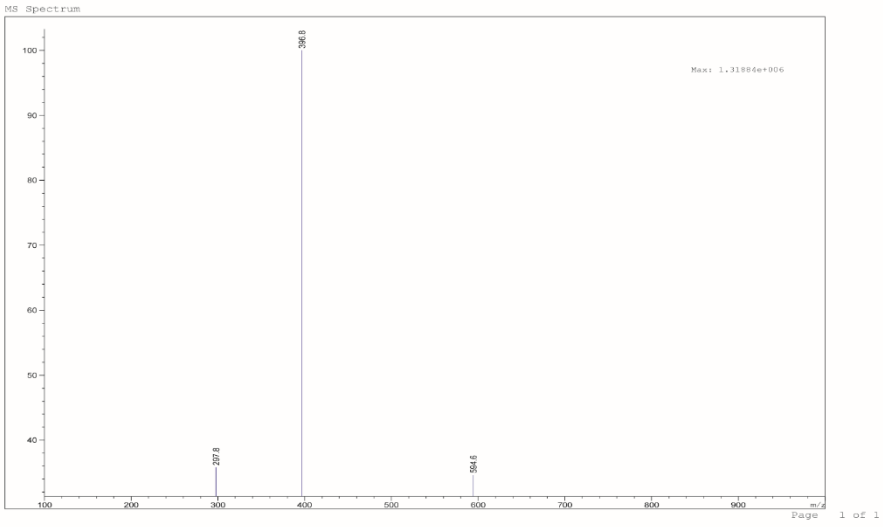
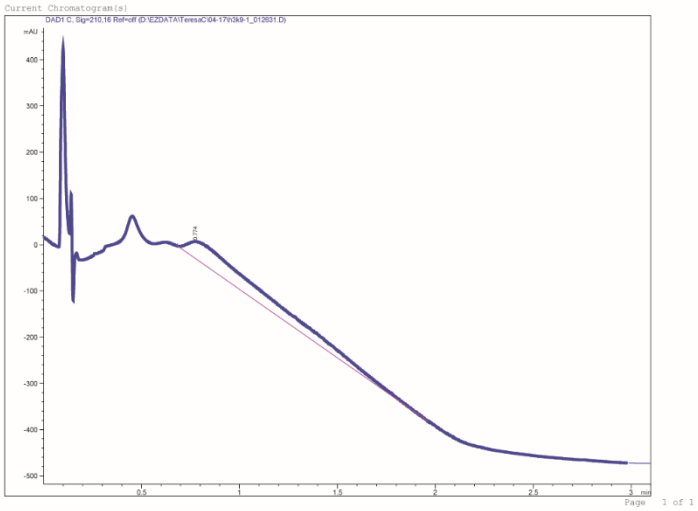


Figure A8.2

LC-MS H3K4ac 10-mer (MW ~ 1186)

A)



H3K9ac_10mer

B)

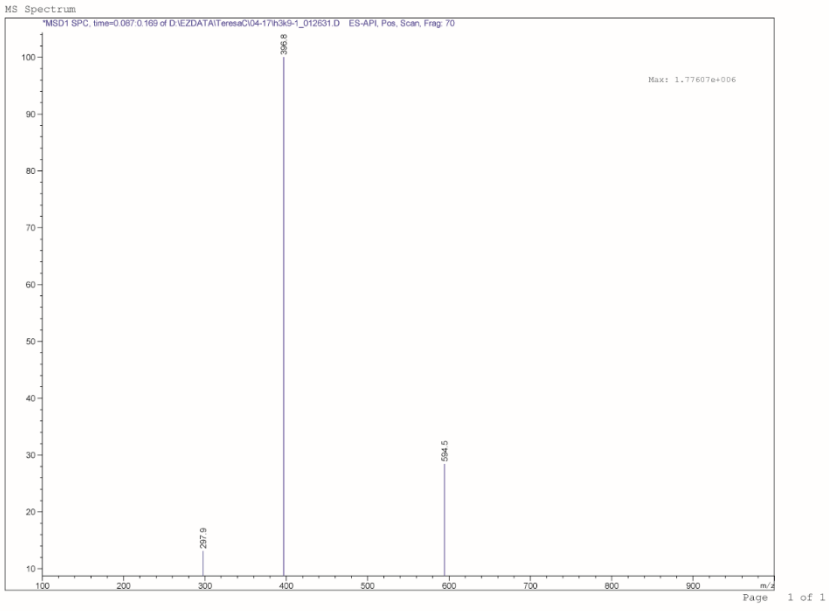
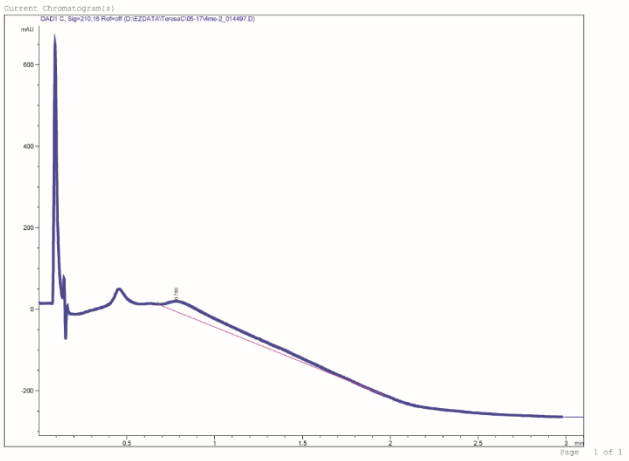


Figure A8.2

LC-MS of peptide H3K9ac 10-mer (MW ~ 1186 Da)

A)



H3K4me_10mer

B)

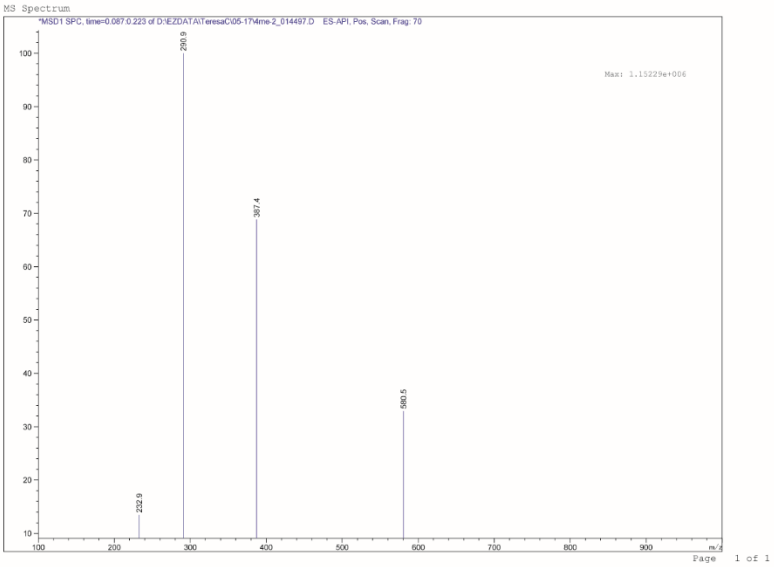


Figure A8. 3

LC-MS H3K4me 10-mer (MW ~1158 Da)

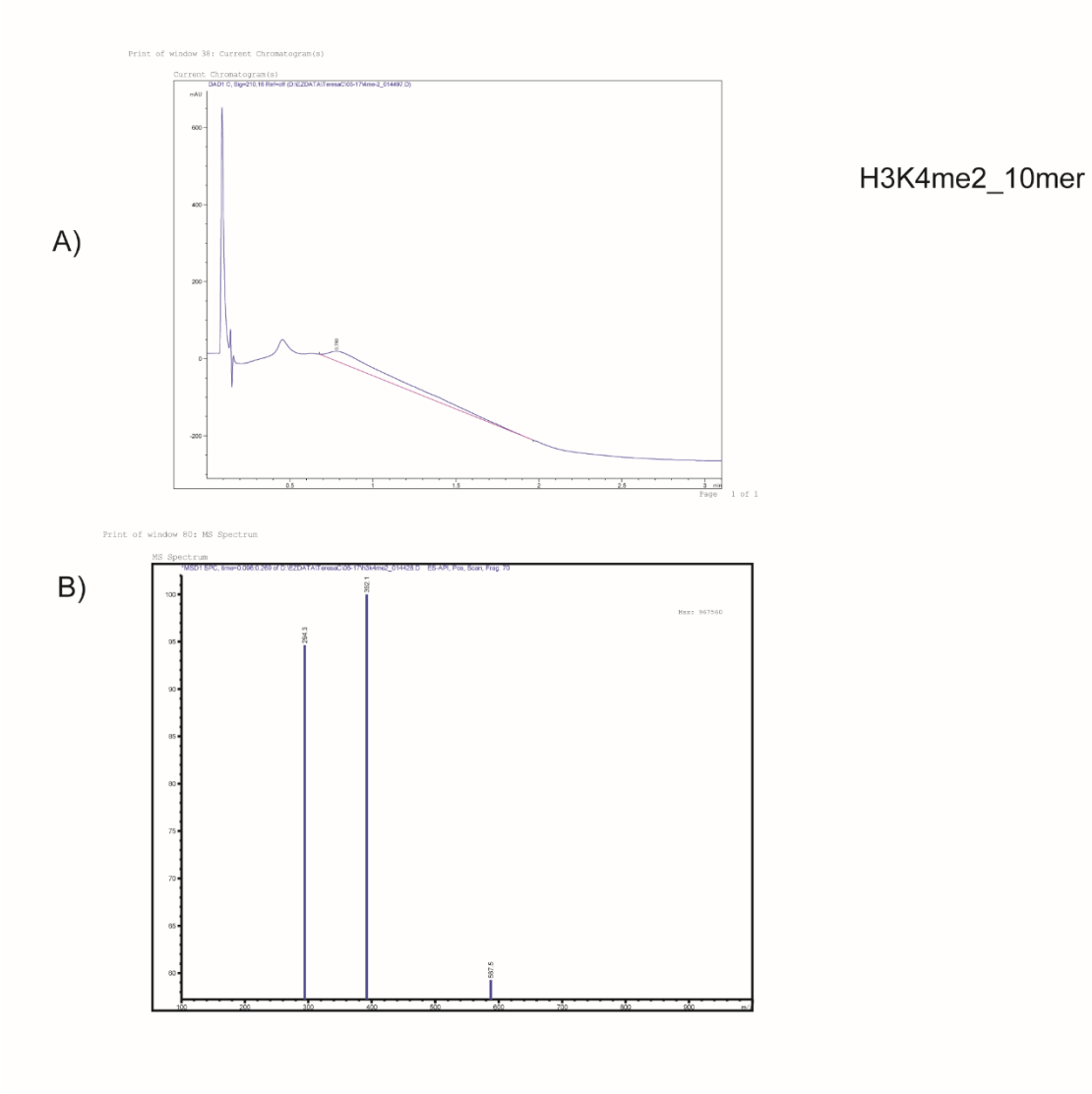


Figure A8.4

LC-MS H3K4me2 10-mer (MW ~1173 Da)

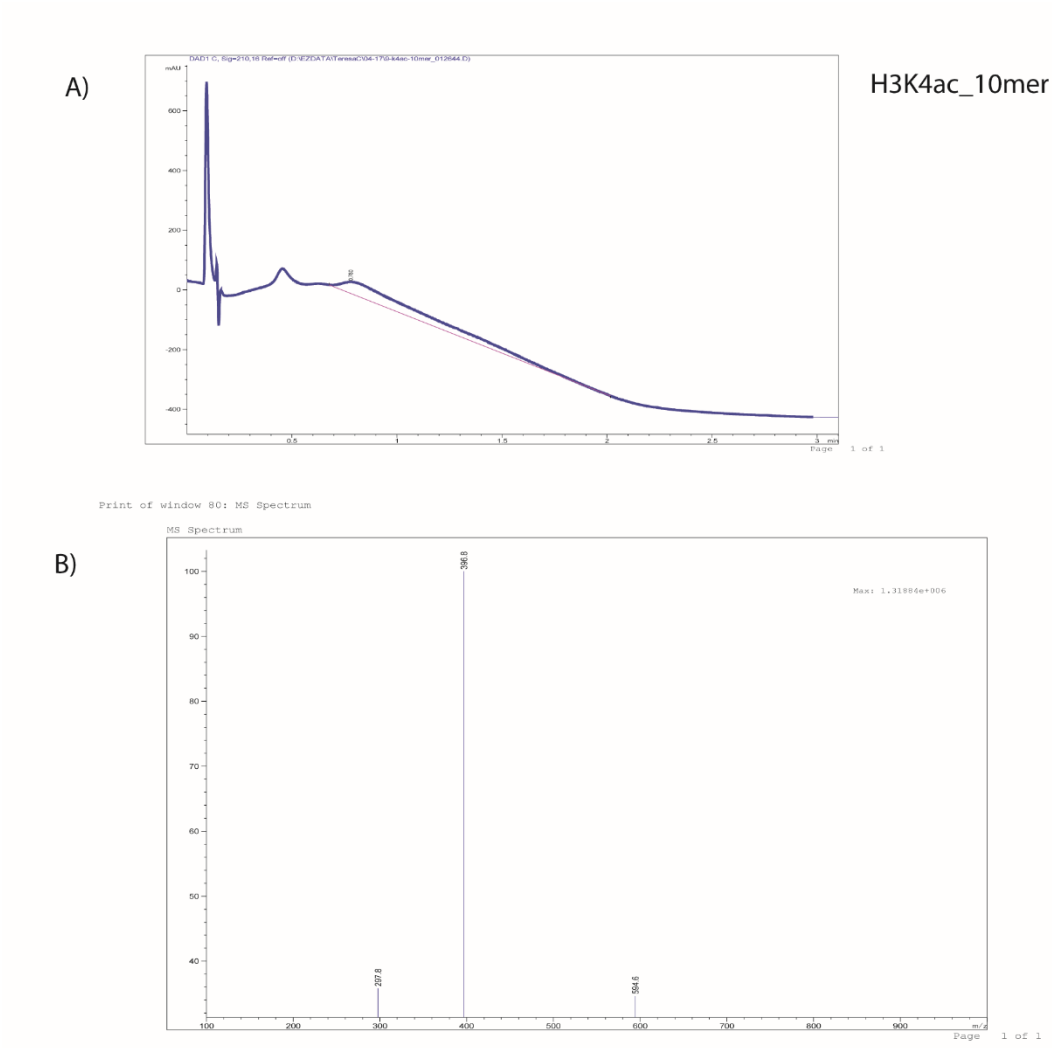


Figure A8.5

LC-MS H3K4me2 10-mer (MW ~1186 Da)

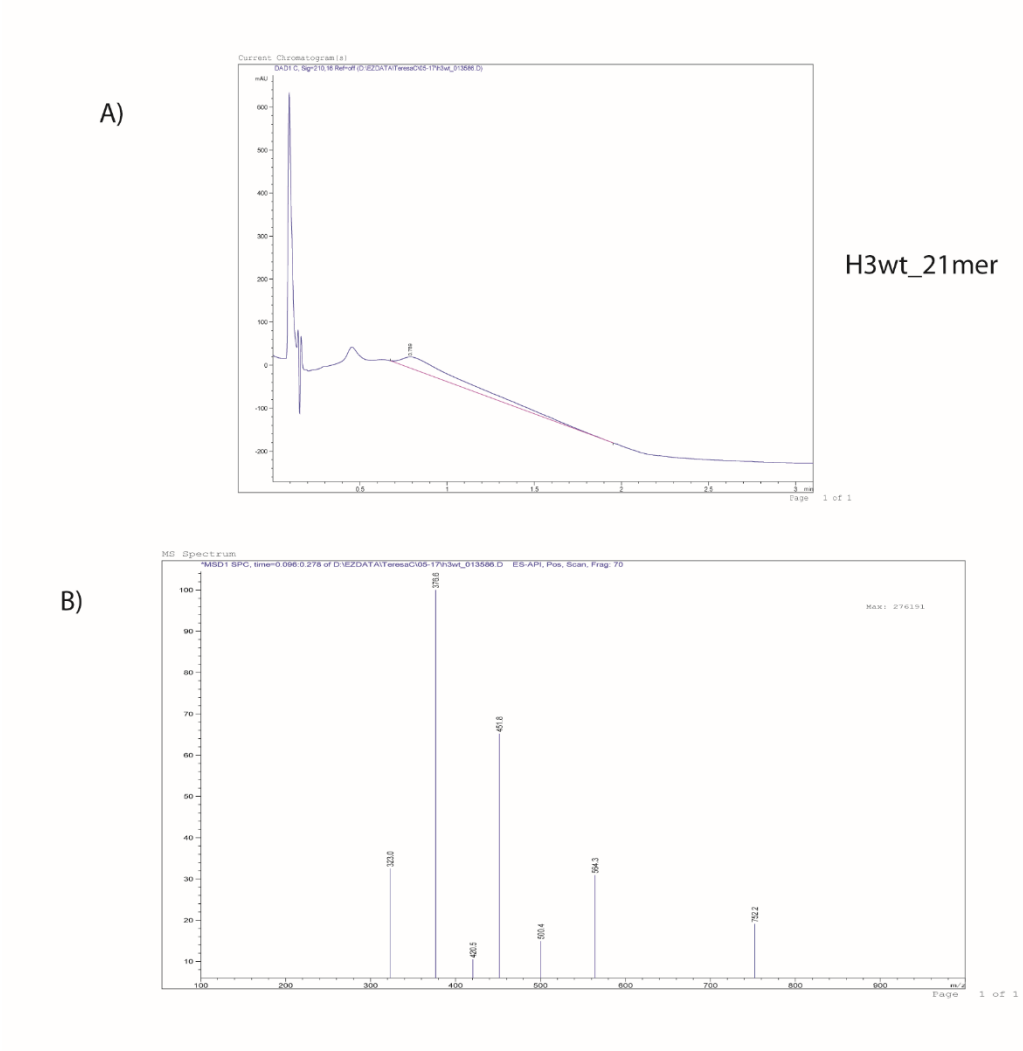
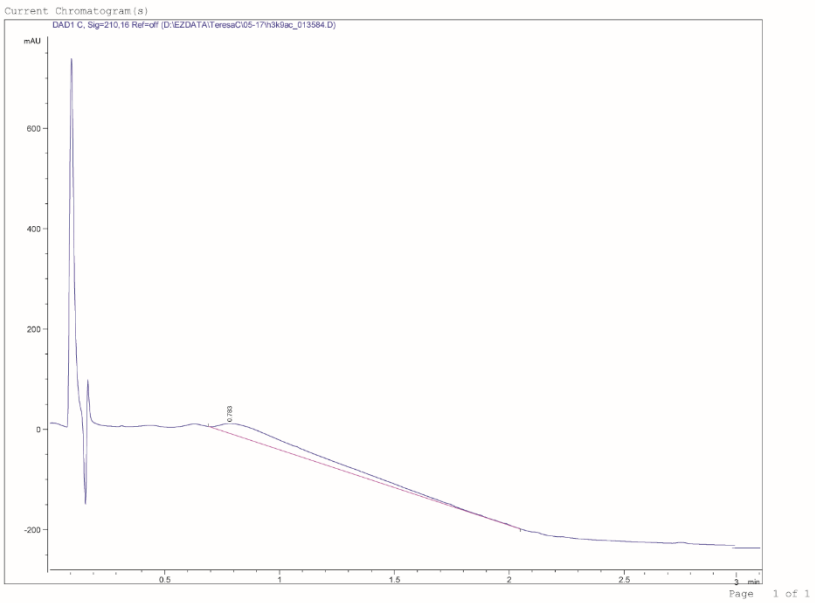


Figure A8.6

LC-MS H3wt 20-mer (MW ~2254 Da)

A)



H3K9ac_21mer

B)

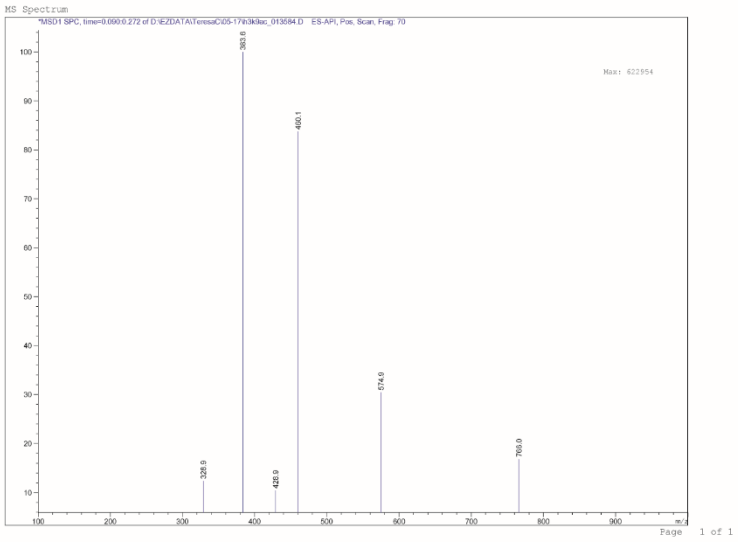
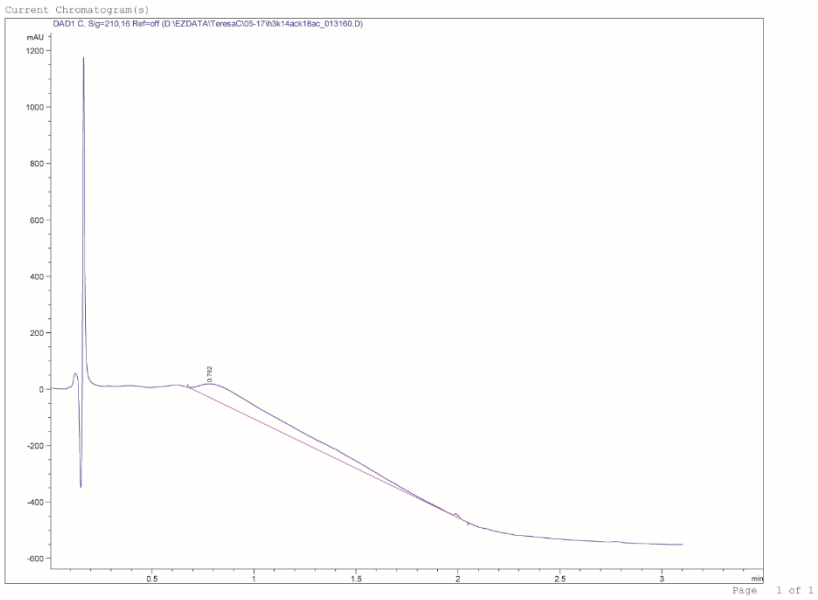


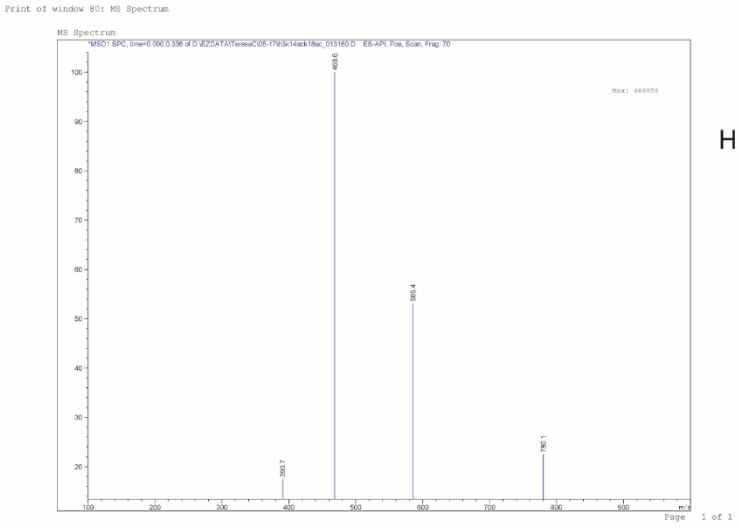
Figure A8.7

LC-MS H3K9ac 20-mer (MW ~2294 Da)

A)



B)



H3K14acK18ac_21mer

Figure A8.8

LC-MS H3K14acK18ac 21-mer (MW ~2334 Da)

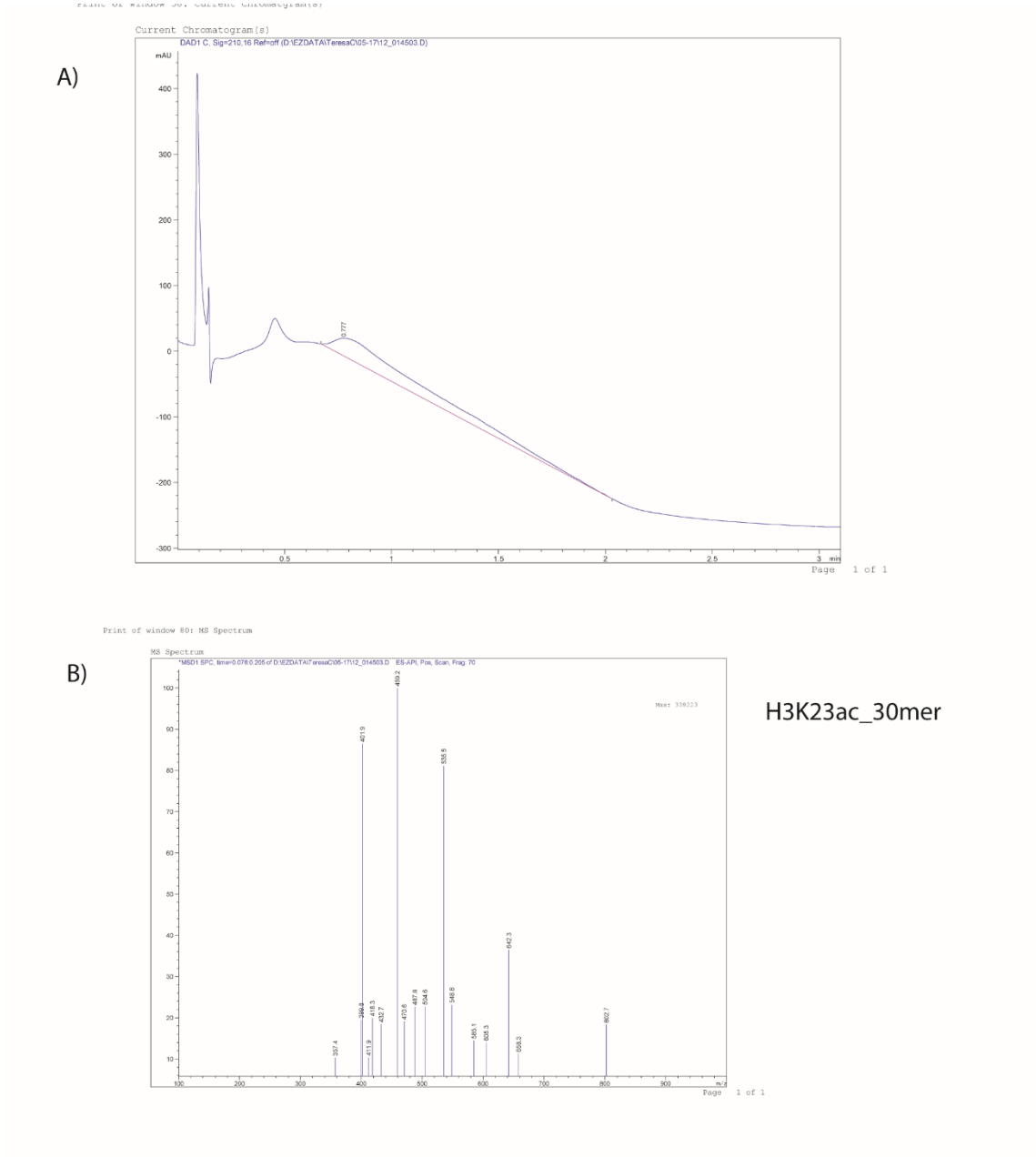


Figure A8.9

LC-MS H3K23ac 30-mer (MW ~3205 Da)

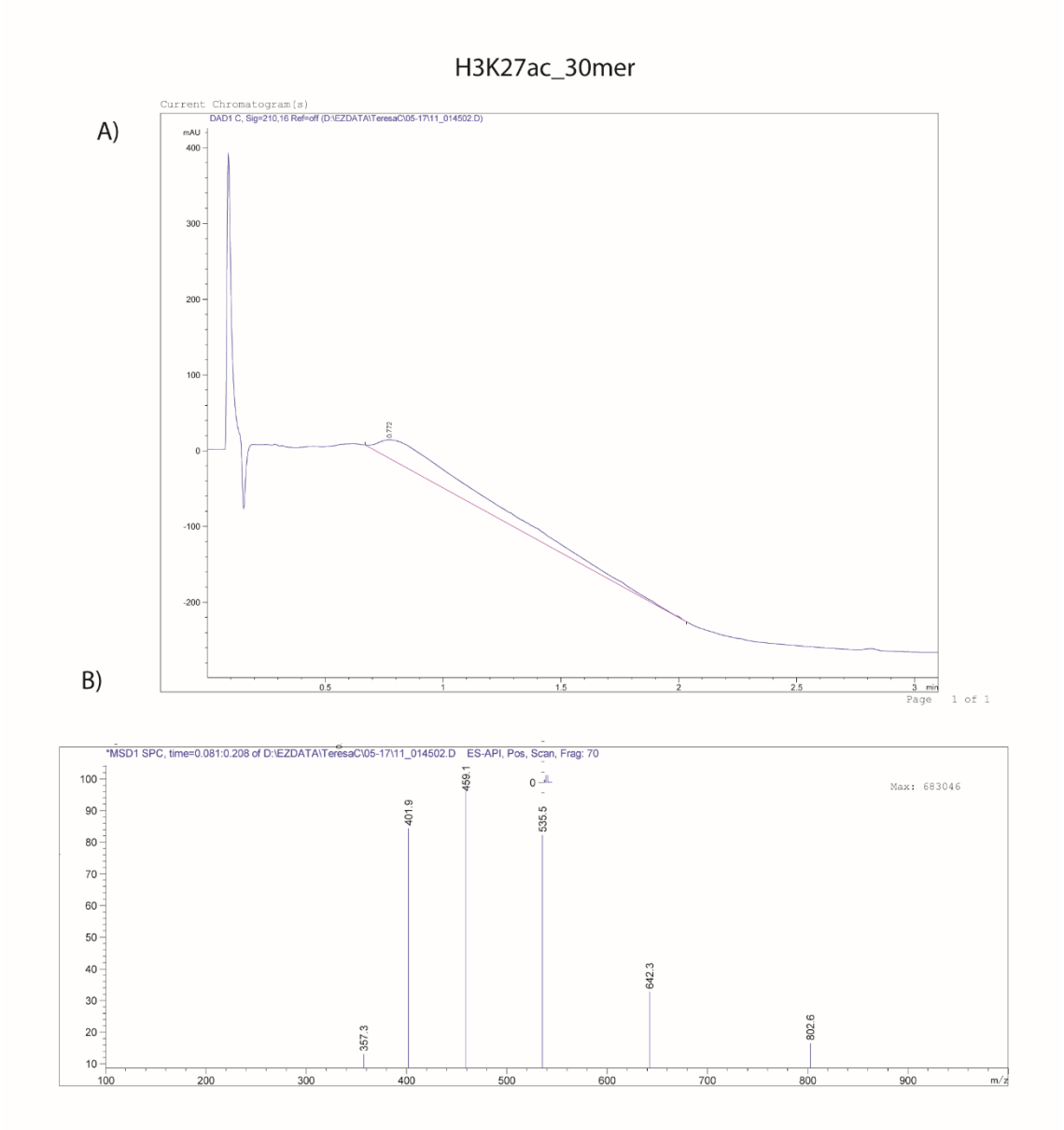


Figure A8.10

LC-MS H3K27ac 30-mer (MW ~3205 Da)

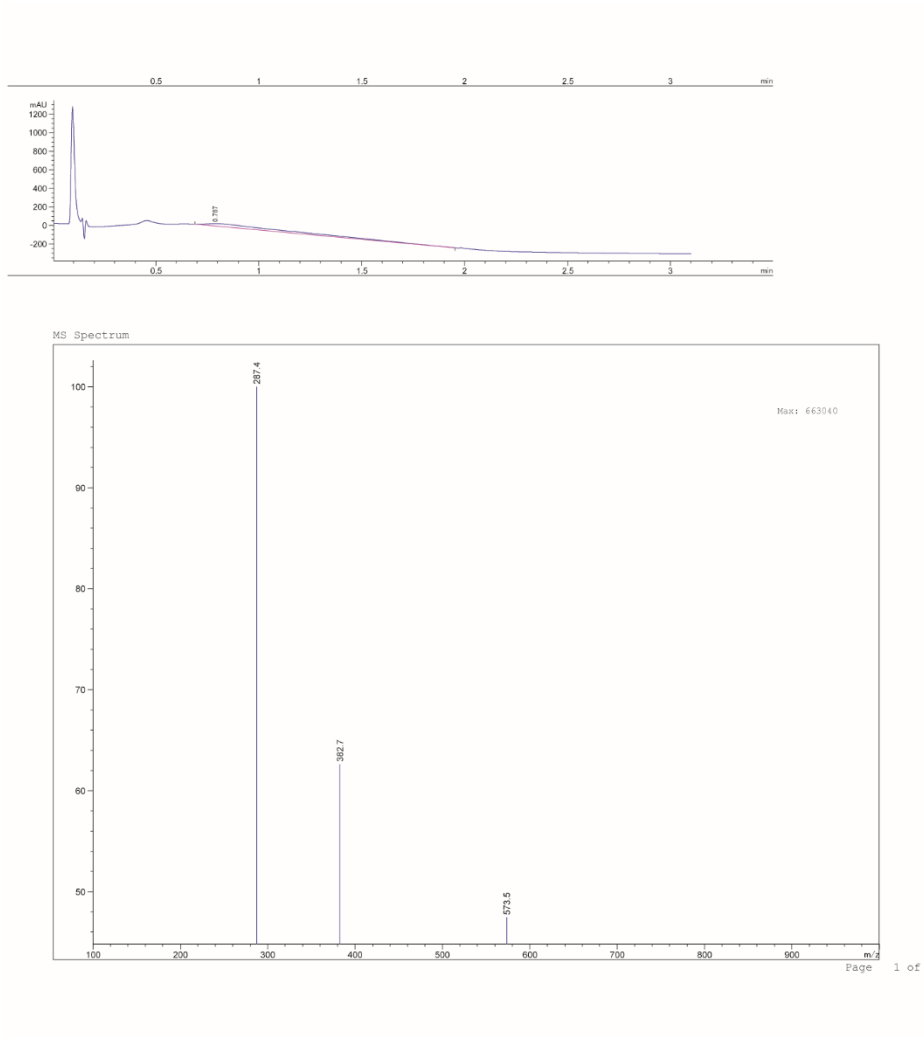


Figure A8.11

LC-MS H3wt 10-mer (MW ~1146 Da)

APPENDIX II

PUBLICATION